

# Pruned skewed Kalman filter and smoother with application to DSGE models

March 2026

Gaygysyz Guljanov<sup>a</sup>, Willi Mutschler<sup>b,c,\*</sup>, Mark Trede<sup>a</sup>

<sup>a</sup>Center for Quantitative Economics, University of Münster, Am Stadtgraben 9, 48143 Münster, Germany.

<sup>b</sup>Department of Economics, University of Tübingen, Mohlstr. 36, 72074 Tübingen, Germany.

<sup>c</sup>Dynare Team, DSGE-net, 56, rue Olivier de Serres, 75015 Paris, France.

---

## Abstract

The *skewed Kalman filter* (SKF) extends the classical *Gaussian Kalman filter* by accommodating asymmetric error distributions in linear state-space models. We introduce a computationally efficient method to mitigate the *curse of increasing skewness dimensions* inherent in the SKF. By analyzing skewness propagation in state-space systems, we develop the *pruned skewed Kalman filter* (PSKF), which eliminates elements in cumulative distribution functions that do not significantly impact asymmetry beyond a specified threshold. Extensive simulations on univariate and multivariate state-space models validate the PSKF's accuracy and efficiency. Additionally, we derive the *skewed Kalman smoother* and its pruned variant, applying them to estimate a New Keynesian DSGE model using US data via standard maximum likelihood and Bayesian MCMC methods. Results strongly favor skewed distributions, particularly for 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

---

---

\*Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—411754673. Declarations of interest: none.

\*\*Replication codes are available at <https://github.com/wmutschl/pruned-skewed-kalman-paper>.

\*Corresponding author [willi@mutschler.eu](mailto:willi@mutschler.eu).

## 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 the closure properties of the normal distribution, see e.g. Azzalini & Capitanio (2014) and Genton (2004) for excellent textbook introductions. More broadly, the CSN represents a special parametrization within the *unified skew normal* (SUN) framework of Arellano-Valle & Azzalini (2006). Notably, it nests both the normal as well as the widely-adopted skew normal distributions of Azzalini (1985), Azzalini & Dalla Valle (1996), and Gupta et al. (2004) 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, most applications remain univariate or impose simplifying model assumptions because of a computational bottleneck, which we term the *curse of increasing skewness dimensions*. The issue stems from the structure of the CSN distribution, whose *probability density function* (pdf) involves multiplying a Gaussian pdf by the ratio of two Gaussian *cumulative distribution functions* (cdf). While low-dimensional Gaussian cdfs are well-understood, the computational cost and numerical error increase sharply with the cdf's dimension. In a Kalman filter setting this difficulty compounds: the number of latent *skewness* components grows with the number of observations and variables, so the dimension of the required cdfs expands throughout the recursion. This manifests the core challenge intrinsic to the skewed Kalman filter, a point recently echoed by Amsler et al. (2021) even for the univariate skew normal distribution.

To address this challenge, our primary contribution is to derive a computationally efficient method to approximate the updating distribution of the skewed Kalman filter by reducing the skewness dimension at each iteration. Our algorithm exploits 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, while low correlation yields near-Gaussian behavior. In the limiting case of zero correlation, the conditional random variable becomes identical to a normally distributed one, causing the *skewed Kalman filter* (SKF) to morph into the *Gaussian Kalman filter* (KF). We operationalize this insight by applying a pruning strategy 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 that do not substantially distort symmetry. This effectively decreases 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 term this algorithm 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. Third, we derive the backward recursions for the *skewed Kalman smoother*. To our knowledge, we are the first to provide these expressions in closed-form and, more importantly, to implement the smoothing steps using our pruning strategy.

We demonstrate 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 exhibit skewness, the pruned skewed Kalman algorithm (i) filters and smooths the unobserved state vector more accurately than the conventional Kalman algorithm, (ii) evaluates the likelihood with only modest overhead relative to the Gaussian Kalman filter, and (iii) delivers precise maximum likelihood estimators for shock parameters in finite samples. In simulation studies where the exact SKF can also be used as a benchmark, the approximation error introduced by our pruning scheme is negligible. Moreover, using the Kullback-Leibler divergence to quantify the distance between exact and pruned SKF distributions, we find that pruning thresholds can be set relatively high (e.g., 1%) while maintaining accuracy even in multivariate settings.

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).<sup>1</sup> 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* (ML) and Bayesian *Markov Chain Monte Carlo* (MCMC) methods, thus showcasing the applicability and versatility of the PSKF across different estimation paradigms. Third, because Chib & Ramamurthy (2014) analyze the same model and dataset under a multivariate Student-t specification for the structural shocks—capturing excess kurtosis but not skewness—adopting this model enables a direct comparison with the related literature. Namely, our estimation results reveal substantial asymmetry in the distributions of productivity and monetary policy shocks. This finding aligns with suggestions by Christiano (2007), Curdia et al. (2014) and Lindé et al. (2016) that skewness may be a salient feature of shock distributions, although those studies do not incorporate skewness in estimation as we do here.

Our presentation and implementation of the pruned skewed Kalman filter and smoother are designed to be highly general, mirroring the simplicity of the standard Gaussian Kalman filtering and smoothing routines. On the modeling side, researchers can retain their linear state-space framework while adding flexibility by assuming a CSN distribution for the error terms in the state transition equation (nesting both the Gaussian and skew normal cases). On the computational side, any estimation workflow—whether Bayesian or Frequentist—that employs Kalman filtering can adopt our method via a drop-in replacement of the underlying Kalman filtering routine. To facilitate widespread use, we provide model-agnostic implementations of the pruned skewed Kalman filter and smoother in Julia, MATLAB, Python, and R.<sup>2</sup> Notably, we have also integrated the pruned skewed Kalman filter and smoother into Dynare 7 (Adjemian et al., 2026)—the leading software for estimating DSGE models—enabling, for the first time, the simulation and estimation of DSGE models with skewed error terms directly in Dynare.

---

<sup>1</sup>In an earlier working paper (Guljanov et al., 2022), we also estimate the multivariate *Dynamic Nelson-Siegel* term structure model of Diebold et al. (2006)—a multivariate vector autoregressive model cast in state-space form—and find strong evidence for skewed error distributions across 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.

<sup>2</sup>Code available at: <https://github.com/gguljanov/pruned-skewed-kalman>.

### *Related literature*

The (closed) skew normal distribution has been used across fields—including asset returns, growth-at-risk, insurance, mental-health and psychiatric measurement, multivariate time series, risk management and stochastic frontier models (Adrian et al., 2019; Cabral et al., 2014; Chen et al., 2003, 2014; Counsell et al., 2011; Eling, 2012; Emvalomatis et al., 2011; Karlsson et al., 2023; Pescheny et al., 2021; Vernic, 2006; Wei et al., 2021; Wolf, 2022; Zhu et al., 2022). By contrast, skewed Kalman filtering is rarely used in practice due to the *curse of increasing skewness dimension*. For linear state-space models, Naveau et al. (2005) and Cabral et al. (2014) adopt CSN assumptions only for the initial state, which keeps the skewness dimension fixed but causes the initial skewness to decay quickly. The same authors alternatively propose to partition the state into linear and skewed components—extended to mixtures by Kim et al. (2014)—but this approach does not generalize to typical multivariate state-space representations and ultimately still faces the *curse of increasing skewness dimension*. Specifying CSN only in the measurement equation (Arellano-Valle et al., 2019) avoids that growth, but is at odds with ample evidence that asymmetry mainly stems from innovations rather than measurement errors. For nonlinear systems, skewed unscented filters (Rezaie & Eidsvik, 2014, 2016) require refitting the updated distribution. Our approach directly addresses the dimensionality problem via pruning and is applicable in all these cases.

Other options—sequential Monte Carlo or Gaussian mixture filters—can handle asymmetry but scale poorly with state dimension; mixtures also have exponentially decaying tails, can be outlier-sensitive, and incur costly reduction steps (Nurminen et al., 2018). Model-specific Bayesian schemes (e.g. Karlsson et al. (2023) for vector autoregressive models) can perform well but reduce portability. We do not claim the PSKF dominates these methods; rather, its drop-in compatibility and computational efficiency make it a practical tool for multivariate applications.

Macroeconomic data that is typically used for DSGE estimation exhibits non-Gaussian features, including fat tails and asymmetry (Fagiolo et al., 2008; Ascari et al., 2015). When data exhibits skewness, the standard linear Gaussian framework forces this asymmetry into the shock process, resulting in model misspecification. Therefore, Christiano (2007) and Lindé et al. (2016) advocate moving beyond Gaussian assumptions based on the skewness and kurtosis properties of residuals. Likewise, Chib & Ramamurthy (2014) and Curdia et al. (2014) demonstrate that linearized DSGE models with Student-t-distributed shocks outperform their Gaussian counterparts. We show that

accounting for skewness yields similar gains. Given this evidence of misspecification, researchers face a choice: maintain the standard Gaussian assumption despite its violations, or extend the framework to accommodate observed patterns. Our PSKF provides a computationally tractable middle ground—it preserves the linear structure while allowing for asymmetric shocks through a simple drop-in replacement of the Kalman filter. Alternative approaches address asymmetries through different mechanisms. Methods like OccBin (Guerrieri & Iacoviello, 2015) handle inequality constraints (e.g., zero lower bound, borrowing constraints) with piecewise linear solutions. Higher-order perturbation methods or global solution techniques capture endogenous skewness from model nonlinearities and can incorporate skewed shocks (Andreasen, 2012). Markov-switching models handle asymmetries from discrete state changes (Lindé et al., 2016). The PSKF complements these approaches by extending the standard linear Gaussian toolkit for cases where asymmetry primarily stems from exogenous shock distributions, offering a computationally tractable alternative. The CSN distribution and our pruning algorithm are also applicable within these alternative frameworks. Closest to us, Grabek et al. (2011) augment a linearized DSGE model with skew normal innovations, but rely on an ad-hoc two-step estimator that can bias results. We instead deploy the (pruned) SKF within standard likelihood-based estimation workflows.

Lastly, our work aligns with a rich literature on asymmetric relationships in macroeconomics. The business cycle asymmetry where recessions are sharper than expansions has a long tradition (Neftçi, 1984; Sichel, 1993). There is renewed interest (Dupraz et al., 2025) in Friedman et al. (1964)’s model that the economy gets *plucked down* by negative shocks from its ceiling—a pattern our left-skewed productivity shocks naturally generate. On the policy side, Ruge-Murcia (2003) and Surico (2007) find evidence of asymmetric central bank preferences—consistent with Fed Chair William McChesney Martin (1955)’s famous metaphor of central banks *taking away the punch bowl* and our finding of right-skewed monetary policy shocks. More recently, Delle Monache et al. (2024) find asymmetry in the conditional distribution of US GDP growth, Karlsson et al. (2023) show that macro skewness is relevant even at the monthly frequency, and Montes-Galdón & Ortega (2022) document positive skewness in Euro Area monetary policy shocks. We share their views on the importance of asymmetric shocks and likewise present evidence for left-skewed productivity and right-skewed monetary policy shocks, though for US quarterly data.

## 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 Online Appendix A), while Section 6 contains the estimation of the linearized New Keynesian DSGE model on US data. Finally, Section 7 concludes.

## 2. Closed skew normal distribution

In this section, we summarize the definition and properties of the CSN distribution. The exposition and notation follow closely González-Farías et al. (2004a), González-Farías et al. (2004b), Grabek et al. (2011) and Rezaie & Eidsvik (2014).

### 2.1. Definition

The CSN distribution arises from a *latent truncation mechanism* acting on a normally distributed population. It is parameterized by five components that govern distinct aspects of the distribution: the *Gaussian backbone parameters*  $\mu$  (location) and  $\Sigma$  (scale), which characterize the observable variables  $W$  before truncation; and the *skewness parameters*  $\nu$  (threshold),  $\Delta$  (latent scale) and  $\Gamma$  (skewness loading), which specify the latent truncation process  $Z$ . Formally, 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  $\Sigma$  is positive semi-definite, the  $q \times q$  covariance matrix  $\Delta$  is positive definite. Let  $\mu$  and  $\nu$  be real vectors of length  $p$  and  $q$ , respectively, and  $\Gamma$  a real  $q \times p$  matrix. Define

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

Then

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

Let the random vector  $X$  have the same distribution as  $W|Z \geq 0$ . Then  $X$  has a CSN distribution

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

with *normal dimension*  $p$ , *skewness dimension*  $q$ , and parameters  $\mu$ ,  $\Sigma$ ,  $\Gamma$ ,  $\nu$  and  $\Delta$ . The moment generating function of  $X$  for any  $t \in \mathbb{R}^p$  is

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

where  $\Phi_q(\cdot; m, S)$  is the cdf of the multivariate normal distribution with expectation vector  $m$  and covariance matrix  $S$ . If the covariance matrix  $\Sigma$  is non-singular, then  $X$  has the probability density function

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

where  $\phi_p(\cdot; \mu, \Sigma)$  is the pdf of a multivariate normal distribution with mean  $\mu$  and covariance matrix  $\Sigma$ . We do not, however, impose non-singularity in general.

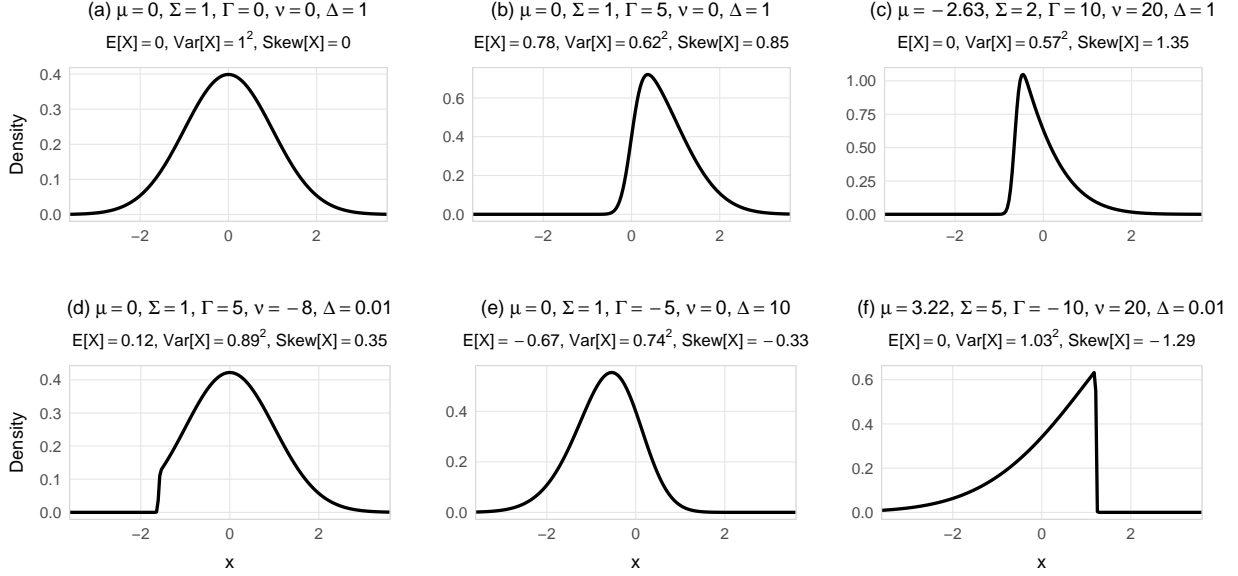
## 2.2. Parameters

Figure 1 illustrates the probability density functions of the univariate CSN distribution under various parameter configurations.<sup>3</sup> Panel (a) shows that  $\Gamma = 0$  yields the standard Gaussian distribution, confirming CSN's role as a proper generalization. Panels (b)–(d) demonstrate right-skewed distributions with positive  $\Gamma$ , while panels (e)–(f) exhibit left-skewness with negative  $\Gamma$ . The figure reveals how the interplay between parameters affects the distribution's shape: comparing panels (b) and (d) shows how negative  $\nu$  values moderate skewness even with identical  $\Gamma$ . Comparing panels (b) and (e) isolates the marginal effect of  $\Delta$ : both share  $|\Gamma| = 5$  and  $\nu = 0$ , but increasing  $\Delta$  from 1 to 10 substantially attenuates skewness, illustrating that larger  $\Delta$  dampens the impact of the skewness parameter  $\Gamma$ . Conversely, small  $\Delta$  values—panels (d) and (f)—sharpen the truncation

---

<sup>3</sup>An interactive web app for exploring the univariate CSN density under arbitrary parameter configurations is available at <https://wmutschl.github.io/csn-univariate-illustration>.

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



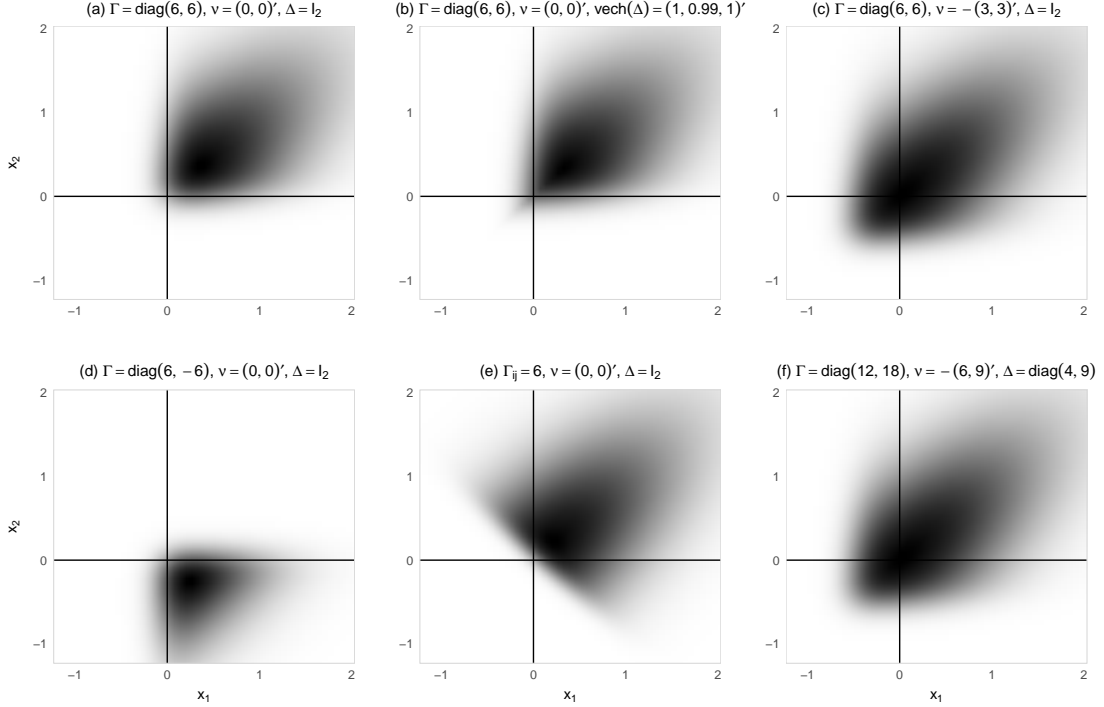
Note: Univariate CSN density  $f_X(x; \mu, \Sigma, \Gamma, \nu, \Delta)$ . Subtitles report analytical moments obtained from the cumulant-generating function  $K_X(t) = \log M_X(t)$ . In panels (c) and (f),  $\mu$  is chosen so that  $E[X] = 0$ .

mechanism, producing more pronounced asymmetry. Importantly, while the univariate skew normal distribution has skewness bounded by  $\pm\sqrt{2}(\pi - 4)/(\pi - 2)^{3/2} \approx \pm 0.995$ , the CSN framework permits substantially higher skewness coefficients, as evidenced by panels (c) and (f) with skewness coefficients exceeding  $\pm 1.29$ .<sup>4</sup> The same panels further illustrate that inverting the analytical cumulant equations from  $K_X(t) = \log M_X(t)$  allows  $\mu$  to be calibrated to any target mean—here chosen to maintain  $E[X] = 0$ .

Similarly, Figure 2 illustrates bivariate CSN distributions under various parameter configurations, all sharing  $\mu = (0, 0)'$  and  $\Sigma$  with unit variances and correlation 0.7. Panel (a) shows same-direction skewness in both margins with independent latent truncation ( $\Delta = I_2$ ). Panel (b) reveals how correlation in the latent truncation mechanism (nonzero off-diagonal in  $\Delta$ ) tilts mass along the main diagonal and strengthens joint tail behavior, with  $\Gamma$  held fixed. Panel (c) demonstrates how negative  $\nu$  shifts the truncation boundary, altering the distribution's shape. Panel (d) exhibits opposing skewness in each margin ( $\Gamma = \text{diag}(6, -6)$ ), creating a distinctive saddle-like contour. Panel (e) shows how nonzero off-diagonal entries in  $\Gamma$  introduce cross-variable skewing, concentrating mass

<sup>4</sup>While we are not aware of a closed-form expression for the maximum skewness achievable by the CSN distribution, Gerber & Pelgrin (2015) provide numerical evidence based on the moment-generating function suggesting the skewness coefficient is bounded by  $\pm 2$ .

Figure 2: Density functions of bivariate CSN distributions with different skewness parameters



*Note:* Density plots of  $CSN_{2,2}(\mu, \Sigma, \Gamma, \nu, \Delta)$  with  $\mu = (0, 0)'$  and  $vech(\Sigma) = (1, 0.7, 1)$ . Panels (c) and (f) illustrate the reparametrization invariance:  $CSN_{2,2}(\mu, \Sigma, \Gamma, \nu, \Delta) = CSN_{2,2}(\mu, \Sigma, A\Gamma, A\nu, A\Delta A')$  with  $A = \text{diag}(2, 3)$ .

along a diagonal ridge. These visualizations underscore that while  $\mu$  and  $\Sigma$  anchor the distribution's location and scale, the skewness parameters  $\Gamma$ ,  $\nu$ , and  $\Delta$  jointly determine both the direction and magnitude of asymmetry through their complex interplay. This flexibility, however, comes with an identifiability issue. The distribution is invariant under the transformation  $A\Gamma$ ,  $A\nu$  and  $A\Delta A'$  for any positive definite diagonal matrix  $A$ , as illustrated by the identical densities in panels (c) and (f) of Figure 2. Consequently, common normalizations fix  $\Delta = I$  and/or constrain  $\nu$  (Arellano-Valle & Azzalini, 2006; Wang et al., 2023) or adopt centered or unified reparametrizations (e.g. scaling  $\Sigma$  to a correlation matrix) to improve numerical stability and parameter interpretability (Arellano-Valle & Azzalini, 2008; Flecher et al., 2009; Käärik et al., 2015). Our pruning algorithm for filtering, smoothing, and likelihood evaluation is agnostic to these conventions—its operations depend only on the implied CSN parameters—and is therefore unaffected by non-identifiability. However, we revisit normalization choices and their practical implications for estimation in Section 5.

Lastly, to contextualize the CSN distribution within the broader skew normal literature, we highlight three important special cases. First,  $CSN_{1,1}(0, 1, \gamma, 0, 1)$  reduces to the univariate stan-

standardized *skew normal distribution*  $SN(0, 1, \gamma)$  introduced by Azzalini (1985)—see panel (b) in Figure 1. Second,  $CSN_{p,1}(\mu, \Sigma, \Gamma, 0, \Delta)$  corresponds to Azzalini & Dalla Valle (1996)’s multivariate extension  $SN_p(\mu, \Sigma, \alpha)$  where  $\alpha = \Delta^{-1/2}\Gamma'$ . Third,  $CSN_{p,q}(\mu, \Sigma, \Gamma, 0, I_q)$  yields the multivariate skew normal distribution  $SN_{p,q}(\mu, \Sigma, \Gamma)$  defined by Gupta et al. (2004)—depicted in panel (e) of Figure 2. While in the univariate case all three collapse to the same skew normal distribution, in multivariate settings the key distinction between  $SN_p(\mu, \Sigma, \alpha)$  and  $SN_{p,q}(\mu, \Sigma, \Gamma)$  lies in their skewness parametrization, i.e. whether to employ a parsimonious  $p$ -dimensional skewness vector or a full  $q \times p$  skewness matrix  $\Gamma$ , trading interpretability for substantially greater flexibility in capturing complex asymmetric patterns and cross-component dependencies. This distinction proves crucial when modeling independent asymmetric shocks—critical for our DSGE application in Section 6. The vector-skewness formulation of Azzalini & Dalla Valle (1996) cannot accommodate independent asymmetric components: even with diagonal  $\Sigma$ , cross-third-order moments remain nonzero because a single latent skewing direction necessarily couples all components and induces co-skewness except in degenerate cases (e.g.  $\alpha$  has at most one nonzero entry).<sup>5</sup> By contrast, both the Gupta et al. (2004) as well as CSN frameworks can preserve independence when the selection acts componentwise; for example, with  $q = p$  and  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ ,  $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_p)$ ,  $\Delta = \text{diag}(\delta_1, \dots, \delta_p)$  and  $\nu = (\nu_1, \dots, \nu_p)'$ . Under these (sufficient) conditions, the latent truncation mechanism factorizes by component, yielding independent marginals. Importantly, these three special cases avoid the identification issue as  $\nu$  and  $\Delta$  are normalized.

### 2.3. Properties

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 a normal distribution. Hence, the CSN can be regarded as a generalization of the Gaussian 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).

---

<sup>5</sup>Let  $X \in \mathbb{R}^p$  be a random variable with unconditional mean vector  $\mu$  and covariance matrix  $\Sigma$ , then we follow Meijer (2005) and define the third central-moment vector as  $\mathcal{M}^{(3)} = \mathbb{E}[(X - \mu) \otimes (X - \mu) \otimes (X - \mu)]$  using the Kronecker product  $\otimes$ . The co-skewness tensor is the standardized cross-third-order central moment:  $skew_{ijk} = [\mathcal{M}^{(3)}]_{ijk} / (\sigma_i \sigma_j \sigma_k)$  with  $\sigma_i = \sqrt{[\text{diag}(\Sigma)]_{ii}}$  such that  $skew_{iii}$  is the usual skewness coefficient.

**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 and  $b \in \mathbb{R}^r$ , then

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

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

In other words, the CSN distribution is closed under linear transformations. If  $F$  is  $p \times p$  square and if both  $F$  and  $\Sigma_x$  have full rank  $p$ , the expressions for  $\Gamma_y$  and  $\Delta_y$  are  $\Gamma_y = \Gamma_x F^{-1}$  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  $\text{rank}(F) = p$  and  $b \in \mathbb{R}^r$ , then

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

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

**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

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

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  $\Gamma_y = 0$  (both  $\nu_y$  and  $\Delta_y$  are irrelevant). Since all elements of the rows in  $\Gamma_z$  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. Given the same partition, we can also compute the marginal distribution of  $X_1$  by applying Property 1:

$$(I \quad 0)X = X_1 \sim CSN_{p_1, q}(\mu_1, \Sigma_{11}, \Gamma_1 + \Gamma_2 \Sigma_{21} \Sigma_{11}^{-1}, \nu, \Delta + \Gamma_2 (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}) \Gamma_2')$$

When  $\Sigma_{11}$  is singular, the formulas extend naturally using the pseudo-inverse, preserving the CSN structure of the conditional and marginal distribution.

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.<sup>6</sup> Particularly, the skewness dimension  $q$  naturally increases each period of the observation window in the filtering algorithm (to be presented in the next section). 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  $\eta_t$  in the transition equation of the state variables originate from the CSN distribution:

$$x_t = Gx_{t-1} + \eta_t, \quad \eta_t \sim CSN_{p, q_\eta}(\mu_\eta, \Sigma_\eta, \Gamma_\eta, \nu_\eta, \Delta_\eta), \quad (2)$$

$$y_t = Fx_t + \varepsilon_t, \quad \varepsilon_t \sim N(\mu_\varepsilon, \Sigma_\varepsilon), \quad (3)$$

where  $x_t$  is the vector of (unobserved) state variables and  $y_t$  the vector of observed variables at equally spaced time points  $t = 1, \dots, T$ . The vector of observation errors  $\varepsilon_t$  is assumed to be

---

<sup>6</sup>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*.

normally distributed and independent of the CSN-distributed state variable shocks  $\eta_t$ . Moreover, we focus on a stable dynamic system, i.e. the characteristic roots of the time-invariant parameter matrix  $G$  are inside the unit circle. In addition, we assume that the distribution of the initial state is anchored at  $CSN(x_{0|0}, \Sigma_{0|0}, \Gamma_{0|0}, \nu_{0|0}, \Delta_{0|0})$ . 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 is 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 may be modeled explicitly (Rezaie & Eidsvik, 2014) or, equivalently, absorbed into the state by adding an auxiliary state to equation (2). In fact, the framework in (2) and (3) is the one most commonly used for economic applications like our analysis 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  $\sigma$ -algebra generated by the observed variables  $\mathcal{F}_t = \sigma(y_t, y_{t-1}, \dots, y_1)$ . 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 before presenting the smoothing equations. Online Appendix C provides the mathematical derivation of the smoothing step—extending Chiplunkar & Huang (2021) from a special case to the general state-space framework—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 (2) in conjunction with closure with respect to linear transformations (Properties 1 and 2) and summation (Property 4) yields the one-step ahead predictive distribution:

$$x_{t|t-1} \sim CSN_{p, q_{t-1}+q_\eta}(\mu_{t|t-1}, \Sigma_{t|t-1}, \Gamma_{t|t-1}, \nu_{t|t-1}, \Delta_{t|t-1}), \quad (4)$$

where

$$\begin{aligned} \mu_{t|t-1} &= G\mu_{t-1|t-1} + \mu_\eta, \quad \Sigma_{t|t-1} = G\Sigma_{t-1|t-1}G' + \Sigma_\eta, & (5) \\ \Gamma_{t|t-1} &= \begin{pmatrix} \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1} \\ \Gamma_\eta\Sigma_\eta\Sigma_{t|t-1}^{-1} \end{pmatrix}, \quad \nu_{t|t-1} = \begin{pmatrix} \nu_{t-1|t-1} \\ \nu_\eta \end{pmatrix}, \quad \Delta_{t|t-1} = \begin{pmatrix} \Delta_{t|t-1}^{11} & \Delta_{t|t-1}^{12} \\ (\Delta_{t|t-1}^{12})' & \Delta_{t|t-1}^{22} \end{pmatrix}, & (6) \end{aligned}$$

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 (3) 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

$$x_{t|t} \sim CSN_{p,q_t}(\mu_{t|t}, \Sigma_{t|t}, \Gamma_{t|t}, \nu_{t|t}, \Delta_{t|t}), \quad (7)$$

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

$$\begin{aligned} \mu_{t|t} &= \mu_{t|t-1} + K_{t-1}^{Gauss}(y_t - F\mu_{t|t-1} - \mu_\varepsilon), \quad \Sigma_{t|t} = \Sigma_{t|t-1} - K_{t-1}^{Gauss}F\Sigma_{t|t-1}, \\ \Gamma_{t|t} &= \Gamma_{t|t-1}, \quad \nu_{t|t} = \nu_{t|t-1} - K_{t-1}^{Skewed}(y_t - F\mu_{t|t-1} - \mu_\varepsilon), \quad \Delta_{t|t} = \Delta_{t|t-1}. & (8) \end{aligned}$$

The updating step consists of two parts, (i) a Gaussian part which updates  $\mu_{t|t}$  and  $\Sigma_{t|t}$  using the *Gaussian Kalman gain*  $K_{t-1}^{Gauss} := \Sigma_{t|t-1}F'(F\Sigma_{t|t-1}F' + \Sigma_\varepsilon)^{-1}$  and (ii) a skewed part which updates the skewness parameters using the *skewed Kalman gain*  $K_{t-1}^{Skewed} := \Gamma_{t|t-1}K_{t-1}^{Gauss}$ . In our setting the only skewness parameter that is updated in the updating step is  $\nu_{t|t-1}$ , the parameters  $\Gamma_{t|t-1}$  and  $\Delta_{t|t-1}$  are not affected because the measurement errors are Gaussian. Again we see that  $\Gamma$  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 (4) and (7) 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).*

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  $y_1, \dots, y_T$ . In the last period the filtered and smoothed distributions obviously coincide. The smoothed distributions for  $t = T - 1, \dots, 1$  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 (2) and (3) with CSN distributed innovations. For ease of notation we define the following matrices:

$$M_t = \Sigma_{t+1|T} \Sigma_{t+1|t}^{-1} G \Sigma_{t|t} \Sigma_{t|T}^{-1} \quad \text{and} \quad N_t = -\Gamma_\eta G + \Gamma_\eta M_t.$$

Further, let  $O_{T-1}, O_{T-2}, \dots$  be a sequence of matrices of increasing row dimensions, such that  $O_{T-1} = N_{T-1}$  and, for  $t = T - 2, T - 3, \dots, 1$ ,

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

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  $t = T - 1, \dots, 1$  are

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

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

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  $t = T - 2, T - 3, \dots, 1$  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 provided in Online Appendix C. 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  $\Sigma, \Gamma, \nu, \Delta$  and a pre-specified pruning threshold  $tol$ .*

1. *Construct and partition the covariance matrix*

$$P = \begin{pmatrix} P_1 & P_2' \\ P_2 & P_4 \end{pmatrix} = \begin{pmatrix} \Sigma & \Sigma \cdot \Gamma' \\ \Gamma \cdot \Sigma & \Delta + \Gamma \cdot \Sigma \cdot \Gamma' \end{pmatrix}. \quad (9)$$

2. Transform  $P$  into a correlation matrix  $R = \begin{pmatrix} R_1 & R'_2 \\ R_2 & R_4 \end{pmatrix}$ .
3. Find the maximum absolute value along each row of  $R_2$ . Save it as vector  $\mathit{max\_val}$ .
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  $(\mathit{max\_val} < \mathit{tol})$ .  
Save as  $\tilde{P}$ .
5. Compute pruned  $\nu$  by removing rows corresponding to  $(\mathit{max\_val} < \mathit{tol})$ .
6. Compute pruned  $\Gamma = \tilde{P}_2 \Sigma^{-1}$ .
7. Compute pruned  $\Delta = \tilde{P}_4 - \Gamma \tilde{P}'_2$ .
8. Return pruned skewness parameters  $\Gamma$ ,  $\nu$ , and  $\Delta$ .

To illustrate the procedure numerically consider the following univariate example:

$$X \sim CSN \left( 0, 1, \begin{pmatrix} 6 \\ 0.1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.1 \\ -0.1 & 1 \end{pmatrix} \right) \quad (10)$$

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

$$R = \begin{pmatrix} 1.0000 & 0.9864 & 0.0995 \\ 0.9864 & 1.0000 & 0.0818 \\ 0.0995 & 0.0818 & 1.0000 \end{pmatrix}.$$

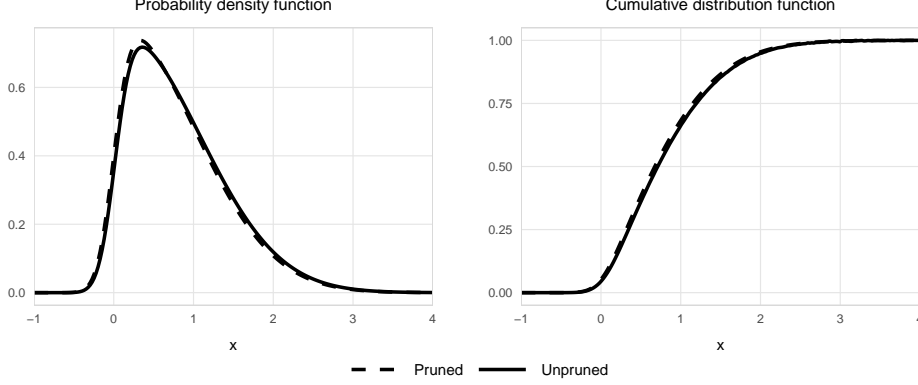
Clearly  $0.9864 > \mathit{tol}$ , but  $0.0995 < \mathit{tol}$ , so we can reduce the skewness dimension by 1. Recomputing the pruned skewness parameters ( $\nu = 0$ ,  $\Gamma = 6 \cdot 1^{-1}$ ,  $\Delta = 37 - 6 \cdot 6$ ), yields the approximating distribution  $CSN(0, 1, 6, 0, 1)$ . Figure 3 depicts the pdf and cdf of the original and the pruned distributions. Despite the large pruning threshold of 0.1, the Kullback-Leibler measure is only 0.00198, indicating that the difference is hardly discernible.<sup>7</sup>

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

---

<sup>7</sup>To provide some intuition: two normal distributions  $N(\mu_1, 1)$  and  $N(\mu_2, 1)$  with identical variances yield a Kullback-Leibler distance of 0.00198 if their expectations differ by only  $\mu_1 - \mu_2 \approx 0.06$ . Such distributions are extremely close. Detecting with probability 0.95 that they are not identical would require more than 4,500 observations in a Gaussian test at a 5% significance level.

Figure 3: Illustration of pruned distribution



*Note:* Solid lines correspond to skewness parameters as given in (10) with two skewness dimensions, dashed lines correspond to the approximating  $CSN(0, 1, 6, 0, 1)$  distribution with one skewness dimension.

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 (6) and (8), with  $\Sigma_{t|t-1}$  as given in (5). Since  $\Gamma_{t-1|t-1}$  appears in the upper row in the prediction step (6), the number of rows increases at each step. For instance, in period  $t = 4$  we would obtain

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

This matrix has dimension  $(4q_{\eta} + q_0) \times p$  where  $p$  is the number of state variables,  $q_{\eta}$  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

$$\Gamma_{t|t} = \begin{pmatrix} \Gamma_{0|0} \Sigma_{0|0} G' \prod_{j=1}^{t-1} L_j \\ \Gamma_{\eta} \Sigma_{\eta} \prod_{j=1}^{t-1} L_j \\ \Gamma_{\eta} \Sigma_{\eta} \prod_{j=2}^{t-1} L_j \\ \vdots \\ \Gamma_{\eta} \Sigma_{\eta} \prod_{j=t}^{t-1} L_j \end{pmatrix} \Sigma_{t|t-1}^{-1}, \quad (11)$$

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  $\Sigma_{t|t-1}$  by  $G$  from the left and by  $\Sigma_{t|t-1}^{-1}$  from the right:

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

As  $t \rightarrow \infty$ , 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 (11) will converge to zero 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 (8) 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 (8). The recursions in (6) imply that the dimension of  $\Delta_{t|t}$  grows each period. The top left element of the partitioned matrix (6) shows that the matrix

$$\begin{aligned} & \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1} - \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1}G\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1} \\ & = \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}^{1/2}(I - \Sigma_{t-1|t-1}^{1/2}G'\Sigma_{t|t-1}^{-1}G\Sigma_{t-1|t-1}^{1/2})\Sigma_{t-1|t-1}^{1/2}\Gamma'_{t-1|t-1} \end{aligned} \quad (12)$$

is added to  $\Delta_{t-1|t-1}$  in each iteration. To show that this additive term is positive definite, we define

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

Since both  $I$  and  $\Sigma_{t|t-1} - G\Sigma_{t-1|t-1}^{1/2}\Sigma_{t-1|t-1}^{1/2}G' = \Sigma_\eta$  (see (5) in the prediction step) are positive definite, so is  $S$  (Horn & Johnson, 2017, Theorem. 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 in equation (12). Since 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 (9), i.e.  $P_2 \equiv \Gamma_{t|t}\Sigma_{t|t}$ . The corresponding correlation

of the  $(i, j)$ -th element  $P_2^{ij}$  is

$$R_2^{ij} = \frac{P_2^{ij}}{\sqrt{\Sigma_{t|t}^{ii}} \sqrt{(\Delta_{t|t} + \Gamma_{t|t} \Sigma_{t|t} \Gamma_{t|t}')^{jj}}}.$$

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,  $(\Delta_{t|t-1} + \Gamma_{t|t-1} \Sigma_{t|t-1} \Gamma_{t|t-1}')^{jj}$  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. To summarize, the algorithm is guaranteed to reduce the skewness dimension after sufficiently many periods.

## 5. A Monte Carlo study

We conduct a comprehensive Monte Carlo study to evaluate the pruned skewed Kalman filter and smoother across three critical dimensions: state estimation accuracy, computational efficiency, and parameter estimation properties. The study employs four carefully designed state-space models as data-generating processes (DGPs), each serving distinct analytical purposes. The first two DGPs focus on algorithmic performance assessing how accurately the filter and smoother recover latent states under various loss functions and quantifying the computational burden of likelihood evaluation—as dimensionality and sample size increase. The latter two DGPs examine parameter estimation challenges under different parametrization strategies. Online Appendix A.1 provides complete specifications of all DGP parameters and detailed simulation results. In what follows, we provide conceptual details on the DGPs and synthesize the key findings.

### 5.1. Data generating processes

The first DGP establishes a baseline univariate system with skewed state transitions. We employ a convenient reparametrization through an auxiliary hyperparameter  $\lambda \in (-1, 1)$ , expressing the skewness parameters as  $\Gamma_\eta = \lambda_\eta \Sigma_\eta^{-1/2}$  and  $\Delta_\eta = (1 - \lambda_\eta^2) I_{n_\eta}$ . This choice elegantly reduces the matrix product  $\Delta_\eta + \Gamma_\eta' \Sigma_\eta \Gamma_\eta$  to the identity matrix, yielding closed-form expressions for the unconditional moments of  $\eta_t$  (also in the multivariate case). Beyond computational convenience,

this parametrization resolves the fundamental identifiability issue inherent in the CSN distribution and facilitates moment-based estimation strategies (Flecher et al., 2009; Käärik et al., 2015).

The second DGP scales up complexity by introducing a multivariate system with four state and three observable variables, directly confronting the *curse of increasing skewness dimension* where computational requirements grow exponentially with sample size and the non-pruned filter becomes computationally prohibitive. While maintaining the auxiliary parametrization for computational tractability, this choice involves no loss of generality—our replication codes include functions to compute first two moments for any valid CSN parametrization.

The third DGP shifts focus to parameter estimation challenges by abandoning the auxiliary parametrization. It features three observables with measurement error and three states with distinct distributional characteristics on the shock processes. The goal is to test the filter’s ability to correctly identify and estimate different shocks—one right-skewed, one symmetric (Gaussian), and one left-skewed—with maximum likelihood. This specification is inspired by the multivariate Dynamic Nelson-Siegel term structure model of Diebold et al. (2006), which we estimated in Guljanov et al. (2022) on US data. For identification, we impose the standard normalization  $\nu_\eta = 0$  and  $\Delta_\eta = I_{n_\eta}$ , then directly estimate  $\mu_\eta$ ,  $\Sigma_\eta$ , and  $\Gamma_\eta$  by maximizing the log-likelihood function.

The fourth DGP operationalizes the DSGE model from Section 6, where structural shocks follow *independent* univariate skew normal distributions. The independence structure permits an elegant representation via Gupta et al. (2004)’s multivariate formulation:  $CSN_{n_\eta, n_\eta}(\mu_\eta, \Sigma_\eta, \Gamma_\eta, 0, I_{n_\eta})$ , where  $\Sigma_\eta$  and  $\Gamma_\eta$  are diagonal matrices. As already hinted in Section 2, this offers three key advantages. First, it ensures unique identification of the structural shock distribution which becomes multivariate skew normal. Second, while univariate skew normal distributions constrain individual shock skewness to approximately  $\pm 0.995$ , the recursive filtering operations generate time-varying skewness parameters for the states and observables that may exceed these bounds, as the CSN framework accommodates more extreme asymmetry through its expanded parameter space. Third, the availability of closed-form univariate moment expressions enables analytical calibration of  $\mu_\eta$  to enforce  $E[\eta_t] = 0$ , thereby preserving the standard zero-mean assumption for structural innovations without numerical optimization. Furthermore, it establishes an analytical one-to-one mapping between interpretable moments (standard deviations and skewness coefficients) and the CSN parameters (diagonal entries of  $\Sigma_\eta$  and  $\Gamma_\eta$ ). Our Dynare implementation adopts this

moment-based parametrization, extending the familiar `stderr` and `corr` syntax with a new `skew` keyword, thereby maintaining continuity with existing DSGE modeling and estimation practices.

### 5.2. Filter initialization and log-likelihood computation

For filter initialization, we adopt two approaches depending on the DGP. For DGPs (1)–(3), we follow Harvey & Phillips (1979) and initialize the state distribution as normal with a diffuse covariance matrix having 10 on the diagonal:  $x_{0|0} \sim N(0, 10I_{n_x})$ . For the DSGE model (DGP(4)), we follow standard practice and compute the initial covariance matrix from the solution to the Lyapunov equation,  $\Sigma_{0|0} = G\Sigma_{0|0}G' + \text{Var}(\eta_t)$ , yielding  $x_{0|0} \sim N(0, \Sigma_{0|0})$ . Our results do not depend on these choices. The log-likelihood function is then computed via the standard predictive decomposition based on the conditional distribution of  $y_t$  given  $y_{t-1}$ , which under our framework follows a CSN distribution:

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

where  $\hat{y}_{t|t-1} = F\mu_{t|t-1} + \mu_\epsilon$  is the predicted value and  $\Omega_{t|t-1} = F\Sigma_{t|t-1}F' + \Sigma_\epsilon$  is the prediction-error covariance matrix from the Gaussian Kalman filter. Notably, when  $\Gamma_{t|t-1} = 0$ , the CSN log-likelihood reduces to the standard Gaussian log-likelihood.

### 5.3. How accurately are states recovered?

To evaluate state estimation accuracy, we conduct extensive Monte Carlo simulations using different loss functions and their corresponding optimal point estimators: the expectation, median, and quantiles of both filtered and smoothed states.<sup>8</sup> We simulate 2400 sample paths for  $x_t$  and  $y_t$  of different lengths 40, 80, and 110 plus a burn-in phase, drawing the shocks  $\eta_t$  from the CSN distribution and the measurement errors  $\varepsilon_t$  from the normal distribution. For each configuration, we compute expected losses by averaging over all replications. In addition, we calculate the expected log-likelihoods (for each pruning tolerance level) as the main ingredients of the Kullback-Leibler divergence. Our results reveal several key findings. First and foremost, the skewed Kalman filter and

---

<sup>8</sup>The loss functions are  $L_1(\tilde{x}_t, x_t) = |\tilde{x}_t - x_t|$ ,  $L_2(\tilde{x}_t, x_t) = (\tilde{x}_t - x_t)^2$  and the asymmetric loss function  $L_q(\tilde{x}_t, x_t) = a|\tilde{x}_t - x_t|$  for  $x_t > \tilde{x}_t$  and  $L_q(\tilde{x}_t, x_t) = b|\tilde{x}_t - x_t|$  for  $x_t \leq \tilde{x}_t$ , where  $x_t$  estimates the true value  $\tilde{x}_t$ . As there is no consensus on multivariate extensions of quantiles (see e.g. Jeong, 2023, footnote 3), we focus only on the quadratic loss function  $L_2$  in the multivariate simulations.

smoother consistently outperform their Gaussian counterparts across all specifications. While the improvements are modest in univariate settings, they become substantial in multivariate cases, where accounting for skewness proves particularly valuable. Second, the pruning algorithm demonstrates remarkable numerical accuracy, achieving equivalence with the non-pruned version up to the twelfth decimal place in univariate cases and the fifth decimal in multivariate settings. The Kullback-Leibler divergence remains below  $2.4 \times 10^{-5}$  for all tested pruning tolerances, confirming virtually no information loss for thresholds up to 1% (see footnote 7 for an intuitive interpretation). Moreover, the choice of pruning threshold has minimal impact on accuracy—measurable differences emerge only in multivariate settings and remain numerically small. Finally, the smoother consistently outperforms the filter (as theory predicts), the performance gains remain stable across sample sizes, and our methods maintain their advantage even under asymmetric loss functions that heavily penalize underestimation.

#### 5.4. *How fast is the computation?*

We benchmark computational performance by measuring the wall-clock time required to compute 1000 log-likelihood evaluations across different sample sizes. The results starkly illustrate the curse of dimensionality plaguing the non-pruned skewed Kalman filter: computation time explodes by a factor of over 200 as the sample size grows from 50 to 250 in univariate settings, while becoming entirely infeasible for multivariate models beyond  $T = 100$ . In sharp contrast, our pruned skewed Kalman filter scales gracefully—its runtime increases by approximately the same factors of roughly 4 (univariate) and 5 (multivariate) as the standard Gaussian Kalman filter when moving from  $T = 50$  to  $T = 250$ . This translates to speedups of 300 to 640 times relative to the non-pruned version, effectively solving the computational bottleneck.

While the Gaussian Kalman filter remains fastest in absolute terms (roughly 15 to 20 times faster than our pruned algorithm), the difference is negligible in practice: at  $T = 250$ , the univariate Gaussian filter requires about 1 millisecond per likelihood evaluation versus 16 to 24 milliseconds for the pruned skewed filter, depending on tolerance. In the multivariate case, the gap is even narrower: 5.2 milliseconds (Gaussian filter) versus 41.5 milliseconds (pruned skewed filter,  $10^{-2}$ ).

Both pale in comparison to particle filters or Sequential Monte Carlo methods, which are typically less accurate and slower by several orders of magnitude. To substantiate this claim, we compare the likelihood evaluations of the PSKF against a standard bootstrap particle filter on the DSGE model

(DGP (4)) with skew normal shocks (see Online Appendix B.5 for full details). As the number of particles  $N$  increases, the log-likelihood values obtained from the particle filter converge *toward* those of the PSKF—the remaining discrepancy in the skew normal case is of the same magnitude as the discrepancy between particle filter and exact Kalman filter log-likelihoods in the Gaussian case (less than 0.5 log-likelihood points in both cases with  $N = 2,000,000$ ). This demonstrates that the approximation error of our pruning heuristic is comparable to the inherent Monte Carlo sampling error of the particle filter itself, introducing no meaningful additional bias. In terms of speed, the PSKF computes the likelihood in under 1 second, whereas the particle filter requires 7 seconds with  $N = 20,000$ , 1 minute with  $N = 200,000$ , and 11 minutes with  $N = 2,000,000$ . Importantly, the particle filter suffers from severe sample impoverishment when measurement errors are small, whereas the PSKF remains stable—a critical advantage since estimation of the Ireland (2004) model drives measurement error variances toward zero even in the standard Gaussian case.

The choice of pruning threshold offers a clear speed-accuracy trade-off: looser tolerances ( $10^{-2}$ ) run about twice as fast as tighter ones ( $10^{-6}$ ). Based on our combined accuracy and timing results, we recommend tolerances of  $10^{-2}$  to  $10^{-4}$  for multivariate applications: 1% when speed is paramount and 0.01% when extra accuracy is desired. In univariate models, tightening to  $10^{-6}$  is feasible at essentially no additional cost relative to  $10^{-4}$  and may be preferred if one wants the most conservative pruning.

### 5.5. *How precise are parameter estimates?*

We assess finite-sample estimation performance using  $R = 1200$  simulated datasets from both the multivariate DGP (3) as well as the DSGE-based DGP (4), estimating their distributional shock parameters via maximum likelihood for increasing sample sizes. For DGP (3), the pruned skewed Kalman filter delivers accurate estimates across all parameter types, with virtually identical performance for pruning thresholds ranging from  $10^{-6}$  to  $10^{-2}$ , demonstrating remarkable robustness. The method correctly identifies both skewed and Gaussian distributions without introducing spurious skewness. Estimation precision follows a clear hierarchy: scale parameters are most accurately estimated, followed by location parameters, while skewness parameters prove somewhat more challenging—reflecting the inherent difficulty of estimating higher-order moments, though the normalized root-mean-squared-error (NRMSE, Atkinson et al., 2019) roughly halves when doubling observations from 100 to 200. For the DSGE-based DGP (4), standard deviation parameters are

estimated with high precision by both the pruned skewed and Gaussian Kalman filters (NRMSE well below 0.04), and the two filters yield nearly identical standard deviation estimates—confirming that skewness in the DGP does not distort second-moment estimation. The skewness parameters are recovered with correct signs and improving precision as sample size grows from  $T = 200$  to  $T = 500$ , with no spurious skewness introduced for the symmetric Gaussian shock.

## 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 applicability under both maximum likelihood and Bayesian estimation frameworks. We emphasize that the purpose is not to compare estimation results across paradigms, but rather to showcase the filter’s and smoother’s versatility and ease of implementation regardless of the chosen estimation approach.

### 6.1. Model equations

The log-linearized model equations are given by:

$$\hat{x}_t = \hat{y}_t - \omega \hat{a}_t, \tag{13}$$

$$\hat{g}_t = \hat{y}_t - \hat{y}_{t-1} + \hat{z}_t, \tag{14}$$

$$\hat{x}_t = \alpha_x \hat{x}_{t-1} + (1 - \alpha_x) E_t \hat{x}_{t+1} - (\hat{r}_t - E_t \hat{\pi}_{t+1}) + (1 - \omega)(1 - \rho_a) \hat{a}_t, \tag{15}$$

$$\hat{\pi}_t = \beta (\alpha_\pi \hat{\pi}_{t-1} + (1 - \alpha_\pi) E_t \hat{\pi}_{t+1}) + \psi \hat{x}_t - \hat{e}_t, \tag{16}$$

$$\hat{r}_t - \hat{r}_{t-1} = \rho_\pi \hat{\pi}_t + \rho_x \hat{x}_t + \rho_g \hat{g}_t + \eta_{r,t}, \tag{17}$$

$$\hat{a}_t = \rho_a \hat{a}_{t-1} + \eta_{a,t}/100, \quad \hat{e}_t = \rho_e \hat{e}_{t-1} + \eta_{e,t}/100, \quad \hat{z}_t = \eta_{z,t}/100. \tag{18}$$

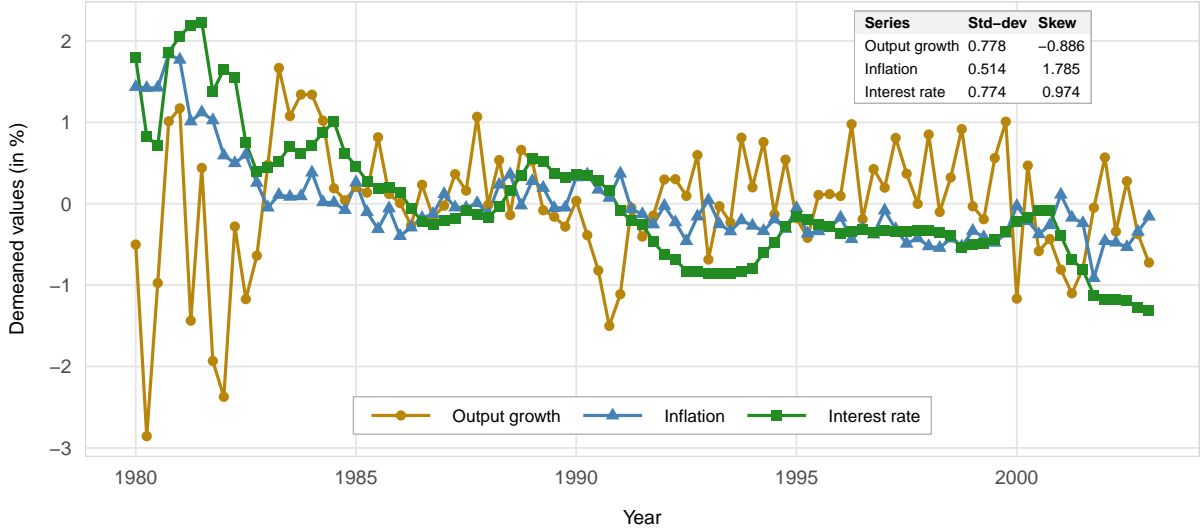
where all hat variables are in log deviations from their non-stochastic steady-state. These equations are based on the optimal behavior of utility-maximizing households and profit-maximizing firms

within a staggered price setting framework. Specifically, the first equation (13) 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.  $\omega$  is a parameter related to the Frisch elasticity of labor and  $\hat{a}_t$  is an autoregressive preference shifter process with persistence parameter  $\rho_a$  and subject to preference shocks  $\eta_{a,t}$ . The second equation (14) defines the growth rate  $\hat{g}_t$  of output subject to productivity shocks  $\eta_{z,t}$ . The third equation (15) 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  $\alpha_x$  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 (16) 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  $\beta$  is the discount factor,  $\psi$  the slope of the curve (influenced by the strength of nominal rigidities) and  $\alpha_\pi$  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  $\rho_e$ . 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 (17) 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.  $\rho_\pi$ ,  $\rho_x$  and  $\rho_g$  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 (18), 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  $\eta_t$ . 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,

Figure 4: Post 1980 US data



Note: Annotation box contains the empirical standard deviation (Std-dev) and skewness (Skew) coefficients.

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 (2) and (3):  $x_t = Gx_{t-1} + R\eta_t$  and  $y_t = Fx_t + \varepsilon_t$ , 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  $\theta$ . We compute these matrices for any given  $\theta$  using Dynare’s first-order perturbation solution algorithm (Villemot, 2011). Note that pre-multiplying  $\eta_t$  by  $R$  in the state transition equation is without loss of generality due to Property 2; thus, we work with the linearly transformed CSN distribution for  $R\eta_t$  in the algorithm.

### 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) (and also in Chib & Ramamurthy (2014)): (1) Demeaned quarterly changes in seasonally adjusted real GDP, converted to per capita values by dividing by the civilian non-institutional 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 US Treasury bill rate provide the measure of the nominal interest rate  $\hat{r}_t$ . Figure 4 displays the three observed series

over 1980Q1–2003:Q1. Because the series are demeaned, sample means are zero by construction; the figure annotations report the empirical standard deviations (Std-dev) and skewness (Skew) coefficients for each series. As previously discussed, while each structural shock is modeled as an independent univariate skew normal process—so its skewness is bounded by  $\pm 0.995$ —this bound need not hold for the observables as they are CSN distributed and can exhibit larger skewness coefficients without any outlier removal or additional data transformations.<sup>9</sup>

Of the ten structural parameters in the model, four are calibrated rather than estimated:  $\beta = 0.99$ ,  $\psi = 0.1$ ,  $\alpha_x = 0$ , and  $\alpha_\pi = 0$ .<sup>10</sup> Hence, our interest centers around the other six model parameters plus the parameters of the distribution of  $\eta_t$ . Rather than assuming conventional multivariate normality, each shock is independently skew normally distributed,  $\eta_{j,t} \sim CSN(\mu_{\eta_j}, \Sigma_{\eta_j}, \Gamma_{\eta_j}, 0, 1)$  for  $j \in \{a, e, z, r\}$ , with  $\mu_{\eta_j}$  set such that  $E[\eta_j] = 0$ . See the discussion of DGP (4) in Section 5 for further implications and motivation of this specification.

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. To illustrate that the PSKF is helpful both for maximum likelihood as well as Bayesian 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 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). Subsequently, we generate 2,000,000

---

<sup>9</sup>For datasets exhibiting extreme asymmetry, alternative heavy-tailed asymmetric families may be more appropriate; for example, Kim & Ruge-Murcia (2019) use the generalized extreme value distribution and estimate a DSGE model using a simulated method of moments approach rather than likelihood-based methods as we do.

<sup>10</sup>The parameters  $\beta$  and  $\psi$  were originally fixed by Ireland (2004) due to lack of identification, while estimates for  $\alpha_x$  and  $\alpha_\pi$  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. We also estimated versions where these parameters are free, but find that the log-likelihood and log-posterior at the mode are actually lower with free  $\alpha_x$  and  $\alpha_\pi$ . To avoid boundary pile-up, multi-modality, and other numerical complications, we therefore set  $\alpha_x$  and  $\alpha_\pi$  to zero in our baseline estimations.

Table 1: Bounds and priors for model and shock parameters

Parameter	BOUNDS		PRIOR		
	Lower	Upper	Type	Mean	Std-dev
$\omega$	0	1	Beta	0.20	0.10
$\rho_\pi$	0	1	Gamma	0.30	0.10
$\rho_g$	0	1	Gamma	0.30	0.10
$\rho_x$	0	1	Gamma	0.25	0.0625
$\rho_a$	0	1	Beta	0.85	0.10
$\rho_e$	0	1	Beta	0.85	0.10
$stderr(\eta_a)$	0	10	InvGamma	$\sqrt{30}$	$\sqrt{30}$
$stderr(\eta_e)$	0	10	InvGamma	$\sqrt{0.08}$	$\sqrt{1}$
$stderr(\eta_z)$	0	10	InvGamma	$\sqrt{5}$	$\sqrt{15}$
$stderr(\eta_r)$	0	10	InvGamma	$\sqrt{0.50}$	$\sqrt{2}$
$skew(\eta_a)$	-0.995	0.995	GenBeta	0	0.40
$skew(\eta_e)$	-0.995	0.995	GenBeta	0	0.40
$skew(\eta_z)$	-0.995	0.995	GenBeta	0	0.40
$skew(\eta_r)$	-0.995	0.995	GenBeta	0	0.40

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

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, while the Online Appendix B provides further computational remarks.

#### 6.4. Estimation results

We caution that Maximum Likelihood (ML) and Bayesian (RWMH) estimates are not directly comparable, as they are obtained under different estimation paradigms; accordingly, we focus first on the ML results and briefly comment on the Bayesian results at the end.

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. Regarding the policy parameters, we observe that allowing for skewness shifts the estimated policy rule being more responsive to movements in the output gap than to output growth as the inflation sensitivity parameter  $\rho_\pi$  is lower, the parameter for the output gap  $\rho_x$  increases, and the one for output growth  $\rho_g$  decreases. While the maximum likelihood estimates for the persistence of the preference shifter,  $\rho_a$ , and the cost-push shock,  $\rho_e$ , are very similar across model variants, the estimate of  $\omega$  is three times higher in the CSN model than in the Gaussian

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
$\omega$	0.0581	0.0685	0.1536	0.1647	0.1273	0.1231	[0.05;0.20]	0.1393	0.1380	[0.05;0.22]
$\rho_\pi$	0.3865	0.2099	0.2876	0.0057	0.4998	0.4938	[0.34;0.66]	0.4675	0.4532	[0.31;0.62]
$\rho_g$	0.3960	0.0612	0.3379	0.0524	0.3626	0.3495	[0.28;0.45]	0.3591	0.3440	[0.28;0.43]
$\rho_x$	0.1654	0.0976	0.2832	0.0114	0.2050	0.1813	[0.12;0.28]	0.2232	0.2013	[0.14;0.30]
$\rho_a$	0.9048	0.0579	0.9170	0.0129	0.9138	0.9189	[0.87;0.96]	0.9213	0.9335	[0.88;0.97]
$\rho_e$	0.9907	0.0130	0.9801	0.0225	0.9105	0.9241	[0.86;0.97]	0.9028	0.9157	[0.85;0.96]
$stderr(\eta_a)$	3.0167	1.5568	2.5232	0.1489	3.2958	3.1659	[1.92;4.64]	3.2292	3.2349	[1.84;4.57]
$stderr(\eta_e)$	0.0248	0.0180	0.0212	0.0462	0.0602	0.0572	[0.05;0.07]	0.0604	0.0572	[0.05;0.07]
$stderr(\eta_z)$	0.8865	0.1245	0.7900	0.0089	0.7884	0.7648	[0.61;0.96]	0.7987	0.7544	[0.62;0.98]
$stderr(\eta_r)$	0.2790	0.0374	0.2838	0.0194	0.2951	0.2796	[0.24;0.35]	0.2920	0.2774	[0.24;0.34]
$skew(\eta_a)$	—	—	-0.1948	0.0176	—	—	—	-0.0823	-0.1972	[-0.47;0.33]
$skew(\eta_e)$	—	—	-0.2140	0.0155	—	—	—	-0.3603	-0.3938	[-0.73;-0.02]
$skew(\eta_z)$	—	—	-0.9953	0.0112	—	—	—	-0.3802	-0.5182	[-0.89;0.15]
$skew(\eta_r)$	—	—	0.8128	0.0193	—	—	—	0.5181	0.6074	[0.23;0.82]
Obj(mode)	1207.56		1215.85		1205.11			1211.47		

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

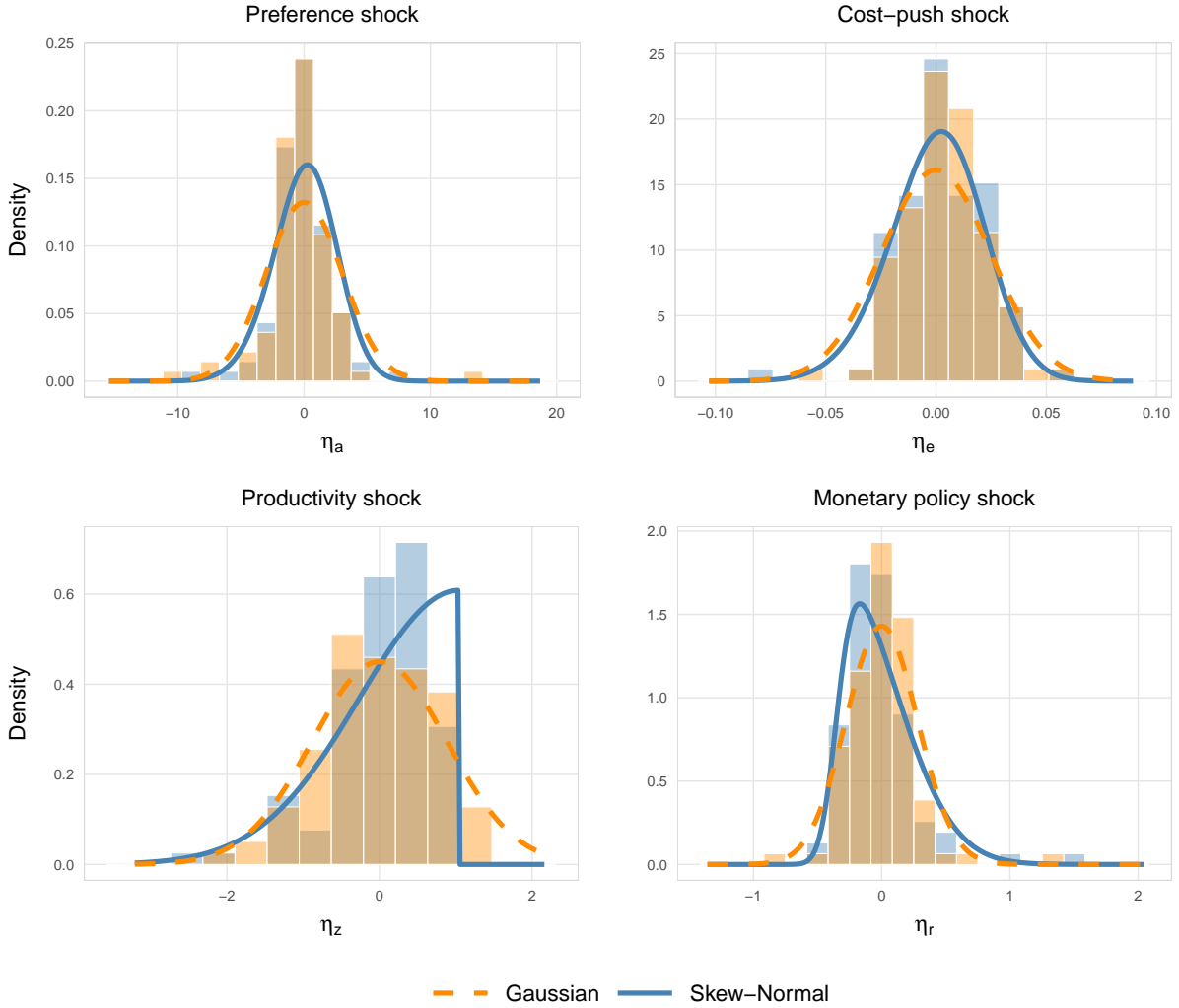
model. This finding has two implications: a higher  $\omega$  implies a more elastic labor supply schedule and the preference shock has a larger impact on the efficient level of output.

Turning to the shock parameters, the estimated standard deviations for the preference ( $\eta_a$ ) and productivity ( $\eta_z$ ) shocks are smaller under the CSN model, while those for the cost-push ( $\eta_e$ ) and monetary policy ( $\eta_r$ ) shocks are very similar across specifications. Notably, all shocks exhibit statistically significant skewness coefficients. The asymmetry is particularly pronounced for the monetary policy shock (0.8171), but even more so for the productivity shock, which at  $-0.995$  reaches its effective theoretical lower bound.<sup>11</sup> In contrast, the preference and (negative) cost-push shocks are estimated to exhibit only low levels of negative skewness.

Figure 5 illustrates these differences by comparing the estimated probability density functions based on the maximum likelihood values, overlaid with histograms of the corresponding smoothed shocks. This provides a direct check of fit: both the orange (Gaussian smoother) as well as blue bars (CSN smoother) align more closely with the estimated skew normal densities (solid line) in the relevant tails, while the estimated Gaussian densities (dashed line) tend to overstate left-tail mass for monetary policy shocks and right-tail mass for productivity shocks, respectively. Taken together,

<sup>11</sup>Obviously, estimates at or near the boundary of the admissible parameter space warrant careful interpretation, as standard asymptotic ML inference may be unreliable in this region (see Online Appendix A.5 for a further discussion).

Figure 5: Estimated probability density functions and smoothed shocks

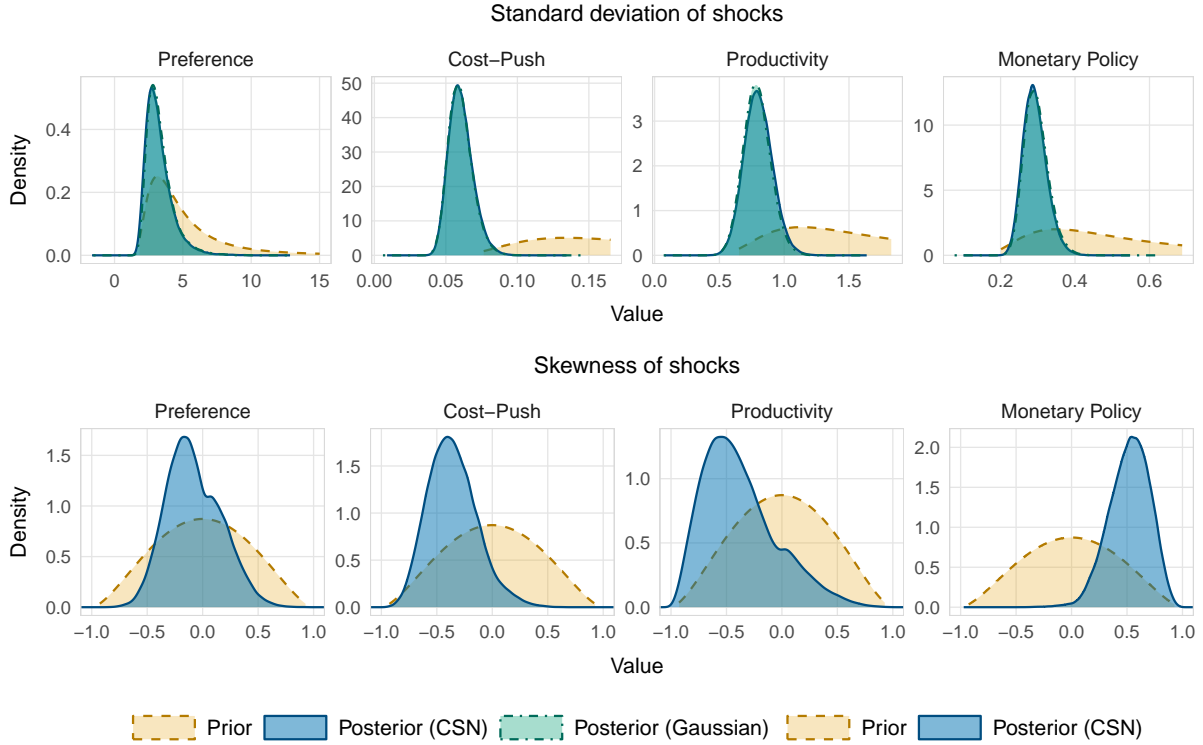


*Note:* Lines plot the model-implied probability density functions of the structural shocks, evaluated at the maximum-likelihood estimates (dashed: Gaussian; solid: skew normal). Histograms show the corresponding smoothed shocks—computed with the Pruned Skewed Kalman Smoother—from the Gaussian (light/orange) and skew normal (dark/blue) specifications; bin widths follow Scott’s rule.

these systematic tail mismatches point to misspecification under Gaussian shocks. The *Pruned Skewed Kalman Filter* provides a tractable remedy—preserving the linear state-space structure while accommodating asymmetric shocks via a drop-in replacement for the Kalman filter.

Finally, the Bayesian estimates largely confirm the insights from maximum likelihood. The log-posterior at the mode is higher for the CSN model (1211.47 vs. 1205.11), and Figure 6 displays the prior and posterior distributions of the shock parameters from the RWMH chains. The top

Figure 6: Prior and posterior distributions of shock parameters



*Note:* Filled areas show kernel density estimates of the prior (orange/dashed) and posterior distributions for the standard deviation and skewness parameters of the structural shocks, estimated under the CSN (blue/solid) and Gaussian (teal/dot-dashed) specifications.

panels show that the standard-deviation posteriors are sharply concentrated and nearly identical across the Gaussian and CSN specifications, confirming that allowing for skewness does not distort inference about second moments. The bottom panels display the skewness posteriors, which are estimated only under the CSN specification. The posterior for the cost-push shock is shifted to the left with a 90% HPD interval of  $[-0.73; -0.02]$  that excludes zero, and the monetary policy shock is clearly positively skewed with a 90% HPD interval of  $[0.23; 0.82]$ . These two posteriors place most of their mass away from zero, providing robust Bayesian evidence against the Gaussian assumption. The preference shock, by contrast, shows little evidence of asymmetry, with a posterior roughly centered around zero. Most notably, the posterior for the productivity shock (mean  $-0.38$ , mode  $-0.52$ , 90% HPD  $[-0.89; 0.15]$ ) lies well inside the admissible range of the skew normal distribution. The prior thus regularizes this estimate and alleviates concerns about the limiting behavior at the parameter boundary of the skew normal distribution.

Table 3: Statistics on recessions in the model

	All			Severe			Mild		
	N	SN	%-dev	N	SN	%-dev	N	SN	%-dev
Number recessions ( $\hat{g}_t \leq -0.5\%$ )	42524	39510	-7.09	14175	13170	-7.09	14175	13170	-7.09
Frequency recessions (in %)	8.50	7.90	-7.09	2.83	2.63	-7.09	2.83	2.63	-7.09
Mean duration (quarters)	3.13	3.07	-1.78	4.44	4.30	-3.23	2.23	2.23	-0.04
Mean output loss (in %)	-2.29	-2.41	5.00	-3.68	-3.92	6.65	-1.18	-1.20	2.18

*Note:* Statistics are based on 500,000 simulated quarters with Gaussian (N) or skew normal (SN) shocks. Calibration is based on column CSN in Table 2. Both variants share the same parameters; Gaussian shocks have skewness set to zero. A severe (mild) recession is a recession associated with a peak-to-trough output loss in the top (bottom) tercile of the distribution. %-dev denotes percentage deviation,  $100(\text{SN}/\text{N} - 1)$ , calculated *before rounding*.

### 6.5. Economic implications

Our primary contribution is methodological; therefore, the estimated model is deliberately stylized and primarily intended for illustrative purposes *conditional on the modeling choice* of attributing asymmetry to shocks. Even so, the estimates yield valuable insights into the nature of business cycles and the role of monetary policy. The right-skewed monetary policy shock implies that large, unexpected tightening episodes—such as sharp increases in the policy rate—are more likely than equally large easing episodes. From 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 (but shorter) 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. This asymmetry can help explain why certain recessionary episodes—such as those associated with oil price spikes (Blanchard & Riggi, 2013), the dot-com bust (Shiller, 2015), or the Gulf War (Federle et al., 2026)—can be more severe than expansions.<sup>12</sup> 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 (N) and the other with skew normal (SN) shocks. The model

<sup>12</sup>The substantial skewness coefficient for productivity confirms the findings of Ruge-Murcia (2017) and seems to suggest that the tails of the skew normal distribution are too thin to capture the large (in absolute value) realizations of this shock. We thank a referee for pointing out that a generalization to the closed skew-t distribution would widen the skewness-kurtosis bound. Our pruning algorithm would be equally applicable to reduce dimensionality.

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 terciles of the implied distribution of peak-to-trough output losses. As Table 3 shows, the CSN model yields slightly fewer total recessions (7.90% vs. 8.50%), which are also somewhat shorter on average (3.07 vs. 3.13 quarters). However, mean output losses are larger (-2.41% vs. -2.29%), driven mainly by deeper severe recessions (-3.92% vs. -3.68%). These patterns align with the model’s assumptions. The strongly left-skewed productivity shocks occasionally trigger sharper downturns, compounded by the right-skewed monetary policy shock, which can deliver unexpected tightening episodes that exacerbate these recessions, 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 do not trigger outsized policy responses or large output drops.<sup>13</sup> Taken together, these findings highlight the importance of modeling the asymmetric nature of shocks rather than defaulting to symmetrical, Gaussian assumptions.

---

<sup>13</sup>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 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.

## 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. 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*. 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 application. In related work, we demonstrate the scalability and applicability of the PSKF to high-dimensional state-space models, namely the Smets & Wouters (2007) model with 20 state variables, embedded in a standard Bayesian MCMC estimation framework (Guljanov, 2024). The evidence of skewness in key macroeconomic shocks aligns with historical patterns: the asymmetric nature of oil price and war shocks (captured by left-skewed productivity) and the tendency for central banks to *take away the punch bowl* more aggressively than to provide stimulus (consistent with right-skewed monetary policy shocks). This challenges the standard framework that assumes symmetric shocks and prompts investigation into the sources of business cycle asymmetry—whether market frictions, policy errors, or structural factors. Importantly, these models can still operate within a linear setting, provided they account for non-normal shock distributions.

Two avenues merit future research. First, distinguishing exogenous skewness (arising from the shock distribution) from endogenous skewness (propagating through asymmetry in the transmission mechanism) would deepen the understanding of business cycle dynamics—a central question for policymakers and for macroprudential interventions in particular. Second, extending the *pruned skewed Kalman filter* to a *pruned skewed Student’s-t filter* would address evidence that both skewness and heavy tails drive business cycle fluctuations. The pruning algorithm applies naturally to this extension.

## **Acknowledgments**

This paper was presented at the Young-Academics Mini-Symposium at the Statistische Woche 2022, the 16th Dynare Conference, the 2023 and 2024 European Meetings of the Econometric Society, the 29th Conference in Computing in Economics and Finance, and the Statistische Woche 2023. The authors thank Dietmar Bauer for both critical and helpful comments as well as kindly sharing his codes on the Mendell-Elston method with us.

## **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, and as a coding assistant for developing data visualizations. These tools were used under human supervision and all authors thoroughly reviewed the final text and code, taking full responsibility for the published content.

## References

- Adjemian, S., Juillard, M., Karamé, F., Mutschler, W., Pfeifer, J., Ratto, M., Villemot, S., & Rion, N. (2026). *Dynare: Reference Manual Version 7*. Dynare Working Papers 87 DSGE-net.
- 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.
- Andreasen, M. M. (2012). On the effects of rare disasters and uncertainty shocks for risk premia in non-linear DSGE models. *Review of Economic Dynamics*, *15*, 295–316. doi:10.1016/j.red.2011.08.001.
- Arellano-Valle, R. B., & Azzalini, A. (2006). On the Unification of Families of Skew-normal Distributions. *Scandinavian Journal of Statistics*, *33*, 561–574. doi:10.1111/j.1467-9469.2006.00503.x.
- Arellano-Valle, R. B., & Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, *99*, 1362–1382. doi:10.1016/j.jmva.2008.01.020.
- Arellano-Valle, R. B., Contreras-Reyes, J. E., Quintero, F. O. L., & Valdebenito, A. (2019). A skew-normal dynamic linear model and Bayesian forecasting. *Computational Statistics*, *34*, 1055–1085. doi:10.1007/s00180-018-0848-1.
- Ascari, G., Fagiolo, G., & Roventini, A. (2015). Fat-Tail Distributions and Business-Cycle Models. *Macroeconomic Dynamics*, *19*, 465–476. doi:10.1017/S1365100513000473.
- Atkinson, T., Richter, A., & Throckmorton, N. (2019). The zero lower bound and estimation accuracy. *Journal of Monetary Economics*, . doi:10.1016/j.jmoneco.2019.06.007.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*, 171–178.
- 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.

- Blanchard, O. J., & Riggi, M. (2013). Why are the 2000s so different from the 1970s? A structural interpretation of changes in the macroeconomic effects of oil prices. *Journal of the European Economic Association*, *11*, 1032–1052. doi:10.1111/jeea.12029.
- 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.
- Christiano, L. J. (2007). Comment. *Journal of Business & Economic Statistics*, *25*, 143–151. doi:10.1198/073500107000000061.
- 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.
- Delle Monache, D., De Polis, A., & Petrella, I. (2024). Modeling and Forecasting Macroeconomic Downside Risk. *Journal of Business & Economic Statistics*, *42*, 1010–1025. doi:10.1080/07350015.2023.2277171.
- 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.
- Dupraz, S., Nakamura, E., & Steinsson, J. (2025). A plucking model of business cycles. *Journal of Monetary Economics*, *152*, 103766. doi:10.1016/j.jmoneco.2025.103766.
- 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.
- Fagiolo, G., Napoletano, M., & Roventini, A. (2008). Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. *Journal of Applied Econometrics*, *23*, 639–669. doi:10.1002/jae.1003.
- Federle, J., Meier, A., Müller, G. J., Mutschler, W., & Schularick, M. (2026). The Price of War. *American Economic Review*, *116*, 791–827. doi:10.1257/aer.20241355.
- Flecher, C., Naveau, P., & Allard, D. (2009). Estimating the closed skew-normal distribution parameters using weighted moments. *Statistics & Probability Letters*, *79*, 1977–1984. doi:10.1016/j.spl.2009.06.004.
- Friedman, M., Conard, J. W., Lary, H. B., & Moore, G. H. (1964). Reports on selected bureau programs. In *The National Bureau Enters Its Forty-Fifth Year* (pp. 7–55). NBER. URL: <http://www.nber.org/chapters/c4453>.
- 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.
- Gerber, M., & Pelgrin, F. (2015). Bayesian Inference for the Multivariate Extended-Skew Normal Distribution. doi:10.48550/ARXIV.1506.05757.
- 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](https://www.nbp.pl/publikacje/materialy_i_studia/101_en.pdf).
- Guerrieri, L., & Iacoviello, M. (2015). OccBin: A toolkit for solving dynamic models with occasionally binding constraints easily. *Journal of Monetary Economics*, *70*, 22–38. doi:10.1016/j.jmoneco.2014.08.005.

- Guljanov, G. (2024). *Estimation of DSGE Models: Skewness Matters*. Ph.D. thesis Universität 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](https://www.wiwi.uni-muenster.de/cqe/sites/cqe/files/CQE_Paper/cqe_wp_101_2022.pdf).
- Gupta, A. K., González-Farías, G., & Domínguez-Molina, J. (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis*, *89*, 181–190. doi:10.1016/S0047-259X(03)00131-3.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, N.J: Princeton University Press.
- Harvey, A. C., & Phillips, G. D. A. (1979). Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances. *Biometrika*, *66*, 49. doi:10.2307/2335241.
- Horn, R. A., & Johnson, C. R. (2017). *Matrix Analysis*. (Second edition, corrected reprint ed.). New York, NY: Cambridge University Press.
- 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.
- Käärik, M., Selart, A., & Käärik, E. (2015). On Parametrization of Multivariate Skew-Normal Distribution. *Communications in Statistics - Theory and Methods*, *44*, 1869–1885. doi:10.1080/03610926.2012.760277.
- Karlsson, S., Mazur, S., & Nguyen, H. (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.
- Kim, J., & Ruge-Murcia, F. (2019). Extreme Events and Optimal Monetary Policy. *International Economic Review*, *60*, 939–963. doi:10.1111/iere.12372.
- 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>.
- Martin, W. M. (1955). Address before the New York Group of the Investment Bankers Association of America. URL: <https://fraser.stlouisfed.org/title/448/item/7800>.
- Meijer, E. (2005). Matrix algebra for higher order moments. *Linear Algebra and its Applications*, *410*, 112–134. doi:10.1016/j.laa.2005.02.040.

- Mendell, N. R., & Elston, R. C. (1974). Multifactorial Qualitative Traits: Genetic Analysis and Prediction of Recurrence Risks. *Biometrics*, *30*, 41. doi:10.2307/2529616.
- Montes-Galdón, C., & Ortega, E. (2022). Skewed SVARs: Tracking the Structural Sources of Macroeconomic Tail Risks. In J. J. Dolado, L. Gambetti, & C. Matthes (Eds.), *Advances in Econometrics* (pp. 177–210). Emerald Publishing Limited. URL: 10.1108/S0731-90532022000044A007.
- 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.
- Neftçi, S. N. (1984). Are Economic Time Series Asymmetric over the Business Cycle? *Journal of Political Economy*, *92*, 307–328. doi:10.1086/261226.
- 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.
- Ruge-Murcia, F. J. (2003). Inflation Targeting under Asymmetric Preferences. *Journal of Money, Credit and Banking*, *35*, 763–785. URL: <https://www.jstor.org/stable/3649827>. doi:<https://www.jstor.org/stable/3649827>. arXiv:3649827.
- Shiller, R. J. (2015). *Irrational Exuberance*. (Revised and expanded third edition ed.). Princeton, New Jersey: Princeton University Press.
- Sichel, D. E. (1993). Business Cycle Asymmetry: A Deeper Look. *Economic Inquiry*, *31*, 224–236. doi:10.1111/j.1465-7295.1993.tb00879.x.
- 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.
- Surico, P. (2007). The Fed’s monetary policy rule and U.S. inflation: The case of asymmetric preferences. *Journal of Economic Dynamics and Control*, *31*, 305–324. doi:10.1016/j.jedc.2005.11.001.

- 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.
- Wang, K., Arellano-Valle, R. B., Azzalini, A., & Genton, M. G. (2023). On the non-identifiability of unified skew-normal distributions. *Stat*, 12, e597. doi:10.1002/sta4.597.
- 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-39237>.
- 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.