Plausibility of big shocks within a linear state space setting with skewness

Koloch, Grzegorz

Warsaw School of Economics

25 January 2016

Plausibility of big shocks within a linear state space setting with skewness*

Grzegorz Koloch†

January 25, 2016

Abstract

In this paper we provide formulae for likelihood function, filtration densities and prediction densities of linear state space model in which shocks are allowed to be skewed. In particular we work with the closed skew normal distribution (csn) introduced in González-Farías et al. (2004), which nests a normal distribution as a special case. Closure of the csn distribution with respect to all necessary transformations in the state space setting is guaranteed by a simple state dimension reduction procedure which does not influence the value of the likelihood function. Presented formulae allow for estimation, filtration and prediction of vector autoregressions and first order perturbations of DSGE models with skewed shocks. This allows to assess asymmetries in shocks, observed data, impulse responses and forecasts confidence intervals. Some of the advantages of using the outlined approach may involve capturing asymmetric inflation risks in central banks forecasts or producing more plausible probabilities of deep but rare recessionary episodes with DSGE/VAR filtration. Exemplary estimation results are provided which show that within a linear setting with skewness frequency of big shocks can be rather plausibly identified.

Keywords: Maximum Likelihood Estimation, State Space Models, Closed Skewed Normal Distribution, DSGE, VAR.

JEL: C51, C13, E32

*This research was financed by the National Science Center grant No. 2012/07/E/HS4/01080.

†Warsaw School of Economics, Institute of Econometrics, Decision Analysis and Support Unit, Al. Niepodleglosci 164, 00-950 Warsaw, Poland.

1 Introduction

A practical motivation derives from an observation emphasised e.g. by Cúrdia, Del Negro and Greenwald (2014), that the magnitude of identified shocks which could lead to significant recessions in usual linear state space settings with normally distributed shocks, within which first order perturbations of DSGE economies are often estimated, are very improbable. For instance, as we show in section 5 using data which contain the recent crisis, in a Schorfheide (2000) economy with shocks fitted with normal distribution, probability of TFP shocks below the 1st percentile is about 2×10^{-4}, which means that, for quarterly data, such a shock should happen once every 1250 years. Cúrdia, Del Negro and Greenwald (2014) propose an approach in which large shocks can occur after replacing normal distribution by a Student's t distribution. They also provide a strong evidence that normality of shocks is counterfactual. Their approach allows for excess kurtosis, but not for for skewness. As they point out, skewness of the distribution of shocks may also be a salient feature of it, and not allowing for skewness may lead to underestimation of the importance of fat tails during recessions.

In this paper we provide a quasi-maximum likelihood estimation procedure for a linear state space model with skewed shocks in the transition equation[1]. As a result, probability of big negative TFP shocks, such as those identified during the recent recession, gets reduced to about 0.5%, which means that such shocks should happen once every 50 years, which seems to be a more plausible frequency.

Since state space form corresponds to some popular macroeconomic tools – VARs and reduced form first order perturbations of DSGE models, we allow for representing data skewness when using such models, within a usual linear framework.

In particular, estimation of VARs and DSGE models with skewed shocks becomes analogical to Kalman filter maximum likelihood estimation and skewness of observables, states and confidence intervals of forecasts can be statistically represented and identified without referring to nonlinearities in the model. Some of the advantages may involve capturing asymmetric inflation risks in central banks forecasts or producing more plausible probabilities of deep but rare contractions with linear VAR

[1] Measurement shocks are assumed to be normally distributed but extension to skewed measurement errors is straightforward.

or DSGE specifications.

In this paper we focus on an econometric part of the agenda, while providing only exemplary results concerning empirical applications, which are left for further study.

Let us consider a following model:

$$
\begin{aligned}
y_t &= Fx_t + Hu_t, \\
x_t &= Ax_{t-1} + B\xi_t, \\
u_t &\sim N(0, \Psi_u), \\
\xi_t &\sim p_\xi(\theta_\xi) \\
x_0 &\sim N(\mu_{x_0}, \Psi_{x_0})
\end{aligned}
$$

for $t \in \mathcal{T} = \{1, 2, ..., T\}$, where $x_t \in \mathrm{R}^p$ and $y_t \in \mathrm{R}^n$ denote states and observables respectively, $\xi_t \in \mathrm{R}^{n_\xi}$ and $u_t \in \mathrm{R}^{n_u}$, $n_\xi, n_u \geq 1$, denote shocks and measurement errors respectively, $\mathrm{R}^{p\times p} \ni A \neq 0$, $B \in \mathrm{R}^{p\times n_\xi}$, $F \in \mathrm{R}^{n\times p}$, $H \in \mathrm{R}^{p\times n_u}$, moreover $\Psi_u \in \mathrm{R}^{n_u\times n_u}$, $\Psi_{x_0} \in \mathrm{R}^{p\times p}$, and $|\Psi_u|, |\Psi_{x_0}| \geq 0$. Finally, $p_\xi(\theta_\xi)$ denotes a probability distribution function of martingale difference shocks ξ_t, which depends on a vector of parameters θ_ξ. Shocks in the measurement equation are assumed to follow a normal distribution for simplicity, they could be skewed as well. The same remark applies to initial states x_0.

A usual assumption is that p_ξ is a multivariate normal distribution, independent across its dimensions. In such a case Kalman filter constitutes an optimal filtration procedure[2], see Simon (2006). If normality assumption is relaxed, Kalman filter remains an optimal *linear* filter. In this paper, we relax normality assumption and assume that elements of ξ_t are independent, but, for some values of θ_ξ, probability density function p_ξ is skewed (asymmetric). In particular, we assume that shocks ξ_t follow a closed skew- normal distribution (*csn* henceforth), which nests the normal distribution as a special case, see González-Farías et al. (2004) or Genton et al. (2004). The *csn* distribution is chosen, because it is closed under almost all transformations imposed on variables in the state space setting[3]. It is not closed, however, under reduced rank linear transformations and we want to allow for rank

[2] In the sense that it minimizes the trace of one-step ahead in-sample forecast errors covariance matrix.
[3] Details are provided in section 2.

deficiency of transition matrix A[4]. This case turns out to be an obstacle in ML estimation since singularity of A precludes propagation of the *csn* distribution through the state space setting. We use a simple state dimension reduction procedure to deal with this issue, so that the *csn* distribution propagates through the state space setting.

To allow for prediction, filtration and estimation, we provide formulae for $p(y_t|Y_{t-1})$, $p(x_t|Y_{t-1})$, $p(x_t|Y_t)$ and for the likelihood function $p(\theta|Y_t)$, where $\theta = (\theta_F, \theta_H, \theta_u, \theta_A, \theta_F, \theta_\xi, \theta_{x_0})$ and $Y_t = \{y_t, y_{t-1}, \ldots, y_1\}$. Although exact formulae are provided, the formula for the likelihood function involves potentially vary large scale normal integrals, practical calculation of which is computationally impossible. To make them operational we factor multivariate integrals into the products of univariate ones, which makes the outlined procedure a quasi-maximum likelihood one. Consequences of this approximation for estimation are unknown in a general case, however our numerical experiments show, that in case of DSGE estimation, this is a very reasonable approximation, since the multivariate normal cumulative distribution functions that appear during calculations are almost perfectly independent across the dimensions.

Remainder of the paper is arranged as follows. In section 2 the closed skew-normal distribution and its basic properties are discussed. Section 3 provides the filter and section 4 the likelihood function. In section 5 we show some estimation results using a DSGE economy of Schorfheide (2000).

2 The skewed normal distribution

In this section we provide the definition of the closed skew-normal distribution, discuss its basic properties and, in particular, discuss its closure under arbitrary linear transformations.

2.1 Definition

Let us denote a density function of a p-dimensional normal distribution with mean[5] μ and a positive definite covariance matrix Σ by $\phi_p(z; \mu, \Sigma)$. Let us also denote a

[4] Which can be the case when the state space form represents a first order perturbation of a DSGE model.

[5] All vectors are column vectors in this paper.

cumulative distribution function of a q-dimensional normal distribution with mean μ and nonnegative definite covariance matrix Σ by $\Phi_q(z;\mu,\Sigma)$. For $q>1$ function Φ_q does not have an analytical expression. Let $r(M)$ denote rank of a matrix M. We will now define the closed skew-normal, possibly singular, distribution by means of the moment generating function (mgf) and then, under nonsingularity of covariance matrix, by means of the probability density function (pdf). For an extensive analysis of this distribution we refer the reader to Genton (2004).

Definition 2.1. (*csn distribution – mgf*) Let $\mu \in \mathrm{R}^p$ and $\vartheta \in \mathrm{R}^q$, $p,q \geq 1$. Let $\Sigma \in \mathrm{R}^{p\times p}$ and $\Delta \in \mathrm{R}^{q\times q}$, $|\Sigma|,|\Delta| \geq 0$, and let $D \in \mathrm{R}^{q\times p}$. We say that random variable z has a (p,q)-dimensional closed skew normal distribution with parameters μ, Σ, D, ϑ and Δ if moment generating function of z, $M_z(t)$, is given by:

$$M_z(t) = \frac{\Phi_q(D\Sigma t;\vartheta,\Delta + D\Sigma D^T)}{\Phi_q(0;\vartheta,\Delta + D\Sigma D^T)} \mathrm{e}^{t^T\mu + \frac{1}{2}t^T\Sigma t}$$

which henceforth will be denoted by:

$$z \sim csn_{p,q}(\mu,\Sigma,D,\vartheta,\Delta)$$

If $|\Sigma| > 0$, a *csn* random variable z obtains a probability density function according to:

Definition 2.2. (*csn distribution – pdf*) If a random variable z follows a (p,q)-dimensional, $p,q \geq 1$, closed skewed normal distribution with parameters μ, Σ, D, ϑ and Δ, where $\mu \in \mathrm{R}^p$, $\vartheta \in \mathrm{R}^q$, $\Sigma \in \mathrm{R}^{p\times p}$, $|\Sigma| > 0$, $\Delta \in \mathrm{R}^{q\times q}$, $|\Delta| \geq 0$ and $D \in \mathrm{R}^{q\times p}$, than probability density function of z is given by:

$$p(z) = \phi_p(z;\mu,\Sigma)\frac{\Phi_q(D(z-\mu);\vartheta,\Delta)}{\Phi_q(0;\vartheta,\Delta + D\Sigma D^T)}$$

Density function (1) defines a (p,q)-dimensional *nonsingular* closed skewed normal distribution in the sense that a random variable has (p,q)-dimensional nonsingular closed skewed normal distribution with parameters μ, Σ, D, ϑ and Δ if and only if its density function for every $z \in \mathrm{R}^p$ equals $p(z)$ in (1). Parameters μ, Σ and D have interpretation of location, scale and skewness parameters respectively. Parameters ϑ and Δ are artificial, but inclusion of them allows for closure of the *csn* distribution under conditioning and taking marginals respectively. The q-dimension in Φ_q is also artificial, but it allows for closure of sums of independent variables and for taking the joint distribution of independent (not necessarily *iid*) variables. When Σ, D and Δ are scalars, they will be denoted respectively by σ, d and δ. For $D = 0$, the *csn* distribution reduces to a p-dimensional normal one. Dimension q can be interpreted as a skewness related degree of freedom in the distribution.

2.2 Properties

This sections discusses basic properties of the *csn* distribution. We will concentrate on three critical issues, which are: closure of the distribution under state-space transformations, conjugate inversion for likelihood derivation and on large scale normal integration (the q-dimension expansion). First, however, all relevant remarks and corollaries are outlined.

2.2.1 Corollaries and remarks.

Remark 2.3. For $p = q = 1$, $\vartheta = 0$ and $\Delta = 1$ the *csn* distribution reduces to the Azzalini skew-normal distribution, see Azzalini (1985), Azzalini (1986).

Such a case will be denoted by:

$$z \sim sn(\mu, \sigma, d)$$

Corollary 2.4. *Let $z \sim sn(\mu, \sigma, d)$, then:*

$$\begin{aligned}
\mathrm{E}(z) &= \mu + \sqrt{\frac{2}{\pi}} \frac{d\sigma}{\sqrt{1 + d^2\sigma}} \\
\mathrm{D}(z) &= \sigma - \frac{2}{\pi} \frac{d^2\sigma^2}{1 + d^2\sigma} \\
\mathrm{E}(z - E(z))^3 &= (2 - \frac{\pi}{2}) \left(\sqrt{\frac{2}{\pi}}\right)^3 \left(\frac{d\sigma}{(1 + \sigma d^2)^{\frac{1}{2}}}\right)^3
\end{aligned}$$

It follows that:

Remark 2.5. Let $z \sim sn(\mu, \sigma, d)$, then $\mathrm{E}(z) = 0$ if and only if $\mu = -\sqrt{\frac{2}{\pi}} \frac{d\sigma}{\sqrt{1+d^2\sigma}}$.

Remark 2.6. For $p \geq 1$, $q = 1$, $\vartheta = 0$ and $\Delta = 1$ the *csn* distribution reduces to the multivariate skew-normal distribution, see Azzalini and Dalla Valle (1996) or Azzalini and Capitano (1999).

Such a case will be denoted by:

$$z \sim sn_p(\mu, \Sigma, d)$$

Corollary 2.7. *Let $z \sim sn_p(\mu, \Sigma, d)$, then:*

$$\begin{aligned} \mathrm{E}(z) &= \mu + \sqrt{\frac{2}{\pi}}\delta \\ \mathrm{D}(z) &= \Sigma + \mu^T\mu + \sqrt{\frac{2}{\pi}}(\mu\delta^T + \delta\mu^T) \end{aligned}$$

where $\delta = \frac{\Sigma d}{\sqrt{1+d^T\Sigma d}}$.

It follows that:

Remark 2.8. Let $z \sim sn_p(\mu, \Sigma, d)$, then $\mathrm{E}(z) = 0$ if and only if $\mu = -\frac{\Sigma d}{\sqrt{1+d^T\Sigma d}}$.

For higher moments of the multivariate skew-normal distribution see Genton et al. (2001).

Corollary 2.9. *Let $z \sim csn_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)$, then:*

$$\mathrm{E}(z) = \mu + \frac{\Phi_q^\star(D\mu; \vartheta, \Delta + D\Sigma D^T)}{\Phi_q(D\mu; \vartheta, \Delta + D\Sigma D^T)}$$

where:

$$\Phi_q^\star(D\mu; \vartheta, \Delta + D\Sigma D^T) = \sum_{i=1}^{p}\sum_{j=1}^{q}(D\Sigma)_{ij}\Phi_q^j(D\mu; \vartheta, \Delta + D\Sigma D^T)e_i$$

for e_i being a p-dimensional unit vector with the i-th entry being equal to 1 and:

$$\Phi_q^j(D\mu; \vartheta, \Delta{+}D\Sigma D^T) = \phi_1((D\mu)_j; \vartheta_j, (\Delta{+}D\Sigma D^T)_{jj})\Phi_{q-1}((D\mu)_{-j}; \vartheta_{-j}, (\Delta{+}D\Sigma D^T)|(D\mu)_j)$$

where, for a generic vector x, x_j denotes the j-th element of x and x_{-j} denotes x without the j-th element.

For higher moments of the closed skew-normal distribution see González-Farías et al. (2004).

Corollary 2.10. *Let $z \sim csn_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)$, $p, q \geq 1$, for parameters as in definition (2.1). Elements of z are independent if and only if matrices Σ and D are diagonal.*

Corollary 2.11. *Let $z \sim csn_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)$, $p, q \geq 1$, for parameters as in definition (2.1). Let also $w \sim N(\mu_w, \Sigma_w)$, $\Sigma_w > 0$, be independent of z, then:*

$$z + w \sim csn_{p,q}(\mu + \mu_w, \Sigma + \Sigma_w, D\Sigma(\Sigma + \Sigma_w)^{-1}, \vartheta, \Delta + (D(I - \Sigma(\Sigma + \Sigma_w)^{-1}))\Sigma D^T)$$

Corollary 2.12. *Let $z \sim csn_{1,q}(\mu, \sigma, d, \vartheta, \delta)$, $q \geq 1$ and for parameters as in definition (2.1), let also $\rho \neq 0$ and $b \in \mathrm{R}$, then:*

$$\rho z + b \sim csn_{1,q}(\rho\mu + b, \rho^2\sigma, \frac{1}{\rho}d, \vartheta, \delta)$$

Corollary 2.13. *Let $z \sim csn_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)$, $p, q \geq 1$, for parameters as in definition (2.1), let also $A \in \mathrm{R}^{p\times p}$, $A \neq 0$, and $b \in \mathrm{R}^p$, then:*

$$Az + b \sim csn_{p,q}(A\mu + b, A\Sigma A^T, D\Sigma A^{-1}, \vartheta, \Delta)$$

Corollary 2.14. *Let $z \sim csn_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)$, $p, q \geq 1$, for parameters as in definition (2.1), let also $A \in \mathrm{R}^{p\times v}$, $p \leq v$ be a full row rank matrix (i.e. of rank v), then:*

$$Az \sim csn_{v,q}(\mu_A, \Sigma_A, D_A, \vartheta_A, \Delta_A)$$

where:

$$\mu_A = \mu A, \quad \Sigma_A = A\Sigma A^T, \quad D_A = D\Sigma A^T \Sigma_A^{-1},$$
$$\vartheta_A = \vartheta, \quad \Delta_A = \Delta + D\Sigma D^T - D\Sigma A^T \Sigma_A^{-1} A\Sigma D^T$$

Corollary 2.15. *Let $z \sim csn_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)$, $p, q \geq 1$, for parameters as in definition (2.1), let also $A \in \mathrm{R}^{v\times p}$, $v > p$ be a full column rank matrix (i.e. of rank v), then:*

$$Az \sim csn_{v,q}(\mu_A, \Sigma_A, D_A, \vartheta_A, \Delta_A)$$

where:

$$\mu_A = \mu A, \quad \Sigma_A = A\Sigma A^T, \quad D_A = D(A^T A)^{-1}, \quad \vartheta_A = \vartheta, \quad \Delta_A = \Delta$$

Notice that corollary (2.14) relates to an isomorphic case or a case in which dimension of Z gets shrinked, whereas corollary (2.18) relates to the case in which dimension of z gets expanded. In the latter case Σ_A is singular and the resulting distribution of Az is called singular. Below we describe how to take joints and sums of independent (not necessarily iid) csn variables.

Corollary 2.16. *Let $z_i \sim csn_{p_i,q_i}(\mu_i, \Sigma_i, D_i, \vartheta_i, \Delta_i)$, $p_i, q_i \geq 1$, $i = 1, 2, ..., n$, for parameters as in definition (2.1), then the joint random variable: $(z_1^T, z_2^T, ..., z_n^T)^T \sim csn_{\sum_{i=1}^n p_i, \sum_{i=1}^n q_i}(\mu^\star, \Sigma^\star, D^\star, \vartheta^\star, \Delta^\star)$, where:*

$$\mu^\star = (\mu_1^T, \mu_2^T, ..., \mu_n^T)^T, \quad \Sigma^\star = \oplus_{i=1}^n \Sigma_i, \quad D^\star = \oplus_{i=1}^n D_i,$$

$$\vartheta^\star = (\vartheta_1^T, \vartheta_2^T, ..., \vartheta_n^T)^T, \quad \Delta^\star = \oplus_{i=1}^n \Delta_i$$

where operator $\oplus$, for arbitrary matrices A and B, is defined as:

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

Corollary 2.17. *Let $z_i \sim csn_{p,q_i}(\mu_i, \Sigma_i, D_i, \vartheta_i, \Delta_i)$, $p, q_i \geq 1$, $i = 1, 2, ..., n$, for parameters as in definition (2.1), then $\sum_{i=1}^n z_i \sim csn_{p, \sum_{i=1}^n q_i}(\mu^\star, \Sigma^\star, D^\star, \vartheta^\star, \Delta^\star)$, where:*

$$\mu^\star = \sum_{i=1}^n \mu_i, \quad \Sigma^\star = \sum_{i=1}^n \Sigma_i, \quad D^\star = (\Sigma_1 D_1^T, ..., \Sigma_n D_n^T)^T (\Sigma^\star)^{-1},$$

$$\vartheta^\star = (\vartheta_1^T, \vartheta_2^T, ..., \vartheta_n^T)^T, \quad \Delta^\star = \Delta^\oplus + D^\oplus \Sigma^\oplus D^\oplus - [\oplus_{i=1}^n D_i \Sigma_i](\Sigma^\star)^{-1} [\oplus_{i=1}^n D_i \Sigma_i]^{-1}$$

for $\Delta^\oplus = \oplus_{i=1}^n \Delta_i$, $D^\oplus = \oplus_{i=1}^n D_i$ and $\Sigma^\oplus = \oplus \Sigma_i$.

The last thing we need to provide is the conjugate Bayesian inverse, which will be used for filtration of states with respect to the likelihood model given by the observation equation.

Corollary 2.18. *Let there be a prior model $z \sim csn_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)$, $p, q \geq 1$, for parameters as in definition (2.1), and a likelihood equation $y = Fz + Hu$, for $F \in \mathrm{R}^{n \times z}$, $H \in \mathrm{R}^{n \times n_u}$ and $u \sim N(0, \Psi_u)$, then posterior model $(z|y)$ is as follows:*

$$(z|y) \sim csn_{n,q}(\mu_{z|y}, \Sigma_{z|y}, D_{z|y}, \vartheta_{z|y}, \Delta_{z|y})$$

where:

$$\mu_{z|y} = \mu + \Sigma F^T (F \Sigma F^T + H \Psi H^T)^{-1} (y - F\mu)$$

$$\Sigma_{z|y} = \Sigma - \Sigma F^T (F \Sigma F^T + H \Psi_u H^T)^{-1} H \Sigma$$

$$D_{z|y} = \left(\begin{pmatrix} D\Sigma \\ 0 \end{pmatrix} - \begin{pmatrix} D\Sigma F^T \\ 0 \end{pmatrix} (F\Sigma F^T + H\Psi_u H^T)^{-1} F\Sigma \right) \Sigma_{z|y}^{-1}$$

$$\vartheta_{z|y} = \begin{pmatrix} \vartheta \\ 0 \end{pmatrix} - \begin{pmatrix} D\Sigma F^T \\ 0 \end{pmatrix} (F\Sigma F^T + H\Psi_u H^T)^{-1} F\Sigma \ (y - F\mu)$$

$$\Delta_{z|y} = \begin{pmatrix} \Delta + D\Sigma D^T & 0 \\ 0 & I \end{pmatrix} - \begin{pmatrix} D\Sigma F^T \\ 0 \end{pmatrix} (F\Sigma F^T + H\Psi_u H^T)^{-1} \begin{pmatrix} D\Sigma F^T \\ 0 \end{pmatrix}^T - D_{z|y} \Sigma_{z|y} D_{z|y}^T$$

2.2.2 Closure and the dimension reduction procedure

Note that in definition 2.1 matrix Σ is allowed to be singular. If Σ is not positive definite, i.e. $|\Sigma| = 0$, resulting distribution is called *singular*. If Σ is positive definite, i.e. $|\Sigma| > 0$, distribution is called *nonsingular*. The *csn* distribution is "closed" in the sense, that it is closed under *full rank* linear transformations[6]. Isomorphic (square full rank) linear transformations transform nonsingular *csn* variables into nonsigular ones and singular variables into singular ones. Full row, but column rank deficient linear transformations (eg. dimension shrinkage/reduction) transform nonsingular *csn* variables into nonsigular ones. Full column, but row rank deficient linear transformations (eg. dimension expansion) transform nonsingular *csn* variables into singular ones, whereas singular variables remain singular. Both singular and nonsingular variables can be transformed into a non-*csn* distributed variables under a rank deficient transformation, hence, the *csn* distribution is not closed under such transformations, which precludes the *csn* distribution from propagation in the state space setting when rank deficient transformations are possible. This fact is negative for maximum likelihood estimation when the transition matrix A in state space equations is singular, which typically is the case in DSGE modeling.

In what follows we discuss two remarks in this respect. First, we provide necessary and sufficient conditions for the *csn* distribution to propagate under arbitrary linear transformations. Second, if these conditions are not satisfied, which is almost always the case, a simple automated dimension reduction procedure is suggested.

Corollary 2.19. *Let $\eta \in \mathcal{R}^m$ be distributed according to a $csn_{m,q}$ for some $m, q \geq 1$ with parameters μ_η, $\Sigma_\eta \geq 0$, D_η, ϑ_η and $\Delta_\eta > 0$. Let $z = G\eta$, $G \in \mathrm{R}^{p\times m}$. Then, z has a csn distribution if and only if G has full row rank or if $Im(G^T) = Im([G^T|w_i])$ for all $i = 1, 2, ..., q$, where $Im(G)$ denotes image (or range) of G and w_i denotes the i-th row D_η.*

Corollary (2.19) states, that for a *csn* variable η, variable $z = G\eta$ has a *csn* distribution if and only if at least one of two conditions apply. The first condition states that row rank of G is full. The second condition requires that rows of D_η

[6] Under full rank transformation we mean full row rank or full column rank transformation and this definition embraces the case when matrix of the transformation is square and represents an isomorphism. When both the row and the column ranks are not full, transformation is called rank deficient.

be linear combinations of rows of G, in other words, that rows of D_η belong to the image of G^T, i.e. to the row space of G. The first condition can be satisfied only for $p \leq m$. The second condition is always satisfied – for any D_η – if G has a full column rank, which can only be the case for $p \geq m$. For a rank deficient operator G, the second condition is a very demanding one, since D_η can be in principle arbitrary.

The correspondence between proposition (2.19) and the state space formulation (1) is the following:

$$\begin{aligned} z &= x_t \\ G &= [A|B] \\ \eta &= (x_{t-1}^T, \xi_t^T)^T \end{aligned}$$

Since $p < m$, according to proposition (2.19), for the *csn* distribution to propagate, we need G to have full row rank. When one works with medium- or large size DSGE models, the reduced form representation matrix A can be, and usually is, rank deficient. Also combining A with B usually results in G which doesn't have a full row rank. Since the following argument applies to full row rank matrices G, we need to reformulate the model so that G has full row rank, but the value of the likelihood function is unaffected. If $G = [A|B]$ is rank deficient, then some of the states x_t are linear combinations of the the remaining ones, which means, that they can be substituted out from the state-space representation using the remaining ones – both in the transition and in the measurement equation. This does not affect the value of the likelihood function and, moreover, this can be done automatically.

Let us denote by $\bar{x}_t$ the (any) maximal linearly independent subset of states from x_t, and by $\tilde{x}_t$ the remaining states. Numerically, we can find a matrix K such that $\tilde{x}_t = K\bar{x}_t$ (for all $t \in \mathcal{T}$). Rearrange $x_t = [\bar{x}_t^T, \tilde{x}_t^T]^T$ and partition model matrices accordingly, so that:

$$y_t = \begin{pmatrix} F_1 & F_2 \end{pmatrix} \begin{pmatrix} \bar{x}_t \\ \tilde{x}_t \end{pmatrix} + Hu_t$$

and:

$$x_t = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \bar{x}_{t-1} \\ \tilde{x}_{t-1} \end{pmatrix} + \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \xi_t$$

Then, first two state space equations in (1) can be rewritten as follows:

$$y_t = \bar{F}\bar{x}_t + Hu_t$$
$$\bar{x}_t = \bar{A}\bar{x}_{t-1} + \bar{B}\xi_t$$

where $\bar{F} = (F_1 + F_2K)$, $\bar{A} = (A_{11} + A_{12}K)$ and $\bar{B} = B_1$. After such a regularization matrix $G = [\bar{A}, \bar{B}]$ has a full row rank. This allows for further steps to apply. Since $\tilde{x}_t = K\bar{x}_t$, we also have full information about the states which were substituted for.

2.2.3 q-dimension expansion

Because Σ is a $p \times p$ matrix and D is a $q \times p$ matrix, corollary (2.10) implies that it is impossible to have $q = 1$ while keeping elements of z independent for $p > 1$, because it has to be the case that $q = p > 1$ in order for D to be diagonal. This is relevant, because the state variable x_t, in every period consists of the *csn* distributed state from the previous period, say x_{t-1}, plus the *csn*-distributed disturbance[7] ξ_t. Corollary 2.17 implies then, that when we add two *csn* variables we have to add their q-dimensions, so that the q-dimension of x_t is the sum of q-dimensions of x_{t-1} and ξ_t, therefore, according to corollary (2.10), contribution of ξ_t to the q-dimension of x_t in every period cannot be squeezed to eg. 1, but must be equal to the size of ξ_t (i.e. n_ξ), hence the q-dimension of x_t quickly increases with t, when the *csn* distribution gets propagates from shocks to states and to observations within the state-space setting. Let us then note, that probability density function (1) of the *csn* distribution and, as a consequence, the likelihood function (see section section 4), they all involve a cumulative probability distribution function of a q-dimensional normal distribution. To our best knowledge there is no efficient way of calculating large scale normal integrals with an arbitrary correlation structure, therefore we work with the following approximation: $\Phi_q(z, \vartheta, \Delta) \approx \prod_{j=1}^{q} \Phi_1(z_j, \vartheta_j, \Delta_{jj})$, which eliminates the curse of dimensionality. Accuracy of this approximation depends on the correlation structure implied by the covariance matrix Δ. In particular, if Δ is diagonal, the result is exact, not approximated. In the numerical experiments with DSGE models we found, that off-diagonal elements of the covariance matrix, which

[7]Both state from the previous period and the disturbance are transformed by the linear transformation A and B respectively, but let us ignore this fact for the present argumentation (or assume this transformations are identities).

appears in the likelihood function (both in the nominator and in the denominator) are close to zero, whereas diagonal elements are significantly bigger. For instance, in the estimation example presented in section 5, mean absolute off-diagonal element of this matrix amounted to 4.88×10^{-4}, whereas mean of diagonal elements was 1.09. In terms of correlation, mean absolute correlation coefficient between dimensions of variables with such a covariance structure equals 4.16×10^{-4}, which is close to zero as for a correlation.

3 The filter

In this section we provide formulae for prediction densities $p(y_t|Y_{t-1})$, $p(x_t|Y_{t-1})$ and for filtration density $p(x_t|Y_t)$. All the densities depend the parameter vector $\theta = (\theta_F, \theta_H, \theta_u, \theta_A, \theta_B, \theta_\xi, \theta_{x_0})$. First, however, we derive unconditional distributions for states and observables.

The state space setting is assumed to be (1) and distribution p_ξ of shocks ξ_t, $t \in \mathcal{T}$, is assumed to be a *csn* distribution:

$$\xi_t \quad \sim \quad csn_{n_\xi,q}(\mu_\xi, \Sigma_\xi, D_\xi, \vartheta_\xi, \Delta_\xi)$$

for $t = 1, 2, ..., T$ with Σ_ξ, D_ξ, and Δ_ξ being diagonal matrices, and $\vartheta_\xi = 0_q$. For D_ξ to be diagonal, it must be the case, that $q = n_\xi$. Remark (2.5) reduces the degrees of freedom in specification of parameters of shocks, since to have $\mathrm{E}(\xi_t) = 0$, one needs to impose: $\mu_\xi = -\sqrt{\frac{2}{\pi}} \frac{d_{\xi_i}\sigma_{\xi_i}}{\sqrt{\delta_{\xi_i} + d^2_{\xi_i}\sigma_{\xi_i}}}$, $i = 1, 2, ..., n_\xi$, where μ_{ξ_i} is the i-th element of μ_ξ, σ_{ξ_i} is the i-the diagonal element of Σ_ξ and $\delta_{\xi_i} = 1$ is the i-th diagonal element of Δ_ξ. In DSGE modeling it usually is the case that $n_\xi < p$ and $r(B) = n_\xi$, which I assume thereafter.

3.1 Unconditional distributions

In what follows we consider two cases for the distribution of state variables. Section (3.1.1) assumes that A is full rank ($r(A) = p$), so that the *csn* distribution propagates through the state space without obstacles. Section (3.1.2) presents the general case in which A can be rank deficient. If $G = [A|B]$ is full (row) rank, regardless whether A is full rank or not, formulae from section (3.1.2) can be applied directly to the system (1) without regularization. If $G = [A|B]$ is (row) rank deficient, in which case A must be rank deficient, regularization described in section (2.3) needs to be exerted on the system before one employs derived formulae. We present section (3.1.1) only because for full rank A unconditional distributions of states and observables can be derived without using joint distribution of states and shocks, which is simpler. Readers interested in the general case can skip to section (3.1.2). Since formulae for distributions of observables are compatible with both cases, they are given separately in section (3.2).

3.1.1 State distribution - full rank case

In this paragraph we assume A is full rank. If $n_\xi < p$, then $B\xi_t$ does not have a density function since matrix $B\Sigma_\xi B^T$ is singular. If $r(B) = n_\xi$, then $B\xi_t$ follows:

$$B\xi_t \sim csn_{p,q}(\mu_B, \Sigma_B, D_B, \vartheta_B, \Delta_B)$$

where:

$$\mu_B = B\mu_\xi, \quad \Sigma_B = B\Sigma_\xi B^T, \quad D_B = D_\xi(B^T B)^{-1}B^T, \quad \vartheta_B = \vartheta_\xi, \quad \Delta_B = \Delta_\xi$$

which is a (p, q)-dimensional singular csn distribution for $q = n_\xi$. This is true for every $t = 1, 2, ..., T$.

Since normal distribution is a special case of the csn distribution, it can be written that:

$$x_0 \sim csn_{p,1}(\mu_{x_0}, \Sigma_{x_0}, D_{x_0}, \vartheta_{x_0}, \Delta_{x_0})$$

for[8]:

$$\mu_{x_0} = \bar{x}, \quad \Sigma_{x_0} = \Psi_x, \quad D_{x_0} = 0, \quad \vartheta_{x_0} = 0, \quad \Delta_{x_0} = 1$$

Generally, i.e. for $t \in \mathcal{T}$, if:

$$x_{t-1} \sim csn_{p,q_{t-1}}(\mu_{x_{t-1}}, \Sigma_{x_{t-1}}, D_{x_{t-1}}, \vartheta_{x_{t-1}}, \Delta_{x_{t-1}})$$

which is true for $t = 1$, i.e. for x_0, with $q_0 = 1$, than Ax_{t-1} follows:

$$Ax_{t-1} \sim csn_{p,q_{t-1}}(\mu_{A,t-1}, \Sigma_{A,t-1}, D_{A,t-1}, \vartheta_{A,t-1}, \Delta_{A,t-1})$$

[8] $\Delta_{x_0} > 0$ can in fact be arbitrary.

where:

$$\mu_{A,t-1} = A\mu_{x,t-1}, \quad \Sigma_{A,t-1} = A\Sigma_{x,t-1}A^T, \quad D_{A,t-1} = D_{x,t-1}\Sigma_{x,t-1}A^T\Sigma_{A,t-1}^{-1}$$

$$\vartheta_{A,t-1} = \vartheta_{x,t-1}, \quad \Delta_{A,t-1} = \Delta_{x,t-1} + D_{x,t-1}\Sigma_{x,t-1}D_{x,t-1}^T +$$

$$-D_{x,t-1}\Sigma_{x,t-1}A^T\Sigma_{A,t-1}^{-1}A\Sigma_{x,t-1}D_{x,t-1}^T$$

which is a (p, q_{t-1})-dimensional *csn* distribution, and, if $|\Sigma_{x_{t-1}}| > 0$, it is non-singular. This is true for $t = 0$ and also, by induction, for every $t \in \mathcal{T}$, because:

$$x_t = Ax_{t-1} + B\xi_t$$

where Ax_{t-1} and $B\xi_t$ are independent random variables, from which it follows, that x_t is distributed according to:

$$x_t \sim csn_{p,q_t}(\mu_{x_t}, \Sigma_{x_t}, D_{x_t}, \vartheta_{x_t}, \Delta_{x_t})$$

for $q_t = q_{t-1} + q$, where, see remark (2.17):

$$\mu_{x,t} = \mu_{A,t} + \mu_B, \quad \Sigma_{x,t} = \Sigma_{A,t} + \Sigma_B, \quad D_{x,t} = \left[\Sigma_{A,t}D_{A,t}^T, \Sigma_B D_B^T\right]^T \Sigma_{x,t}^{-1}$$

$$\vartheta_{A,t} = \left[\vartheta_{A,t}^T, \vartheta_{B,t}^T\right]^T, \quad \Delta_{x,t} = \Delta_{A,t} \otimes \Delta_B + (D_{A,t} \otimes D_B)(\Sigma_{A,t} \otimes \Sigma_B)(D_{A,t} \otimes D_B)^T +$$

$$-(D_{A,t}\Sigma_{A,t} \oplus D_B\Sigma_B)(\Sigma_{x,t})^{-1}(\Sigma_{A,t}D_{A,t}^T \oplus \Sigma_B D_B^T)$$

Since $q_0 = 1$ and $q_t = q_{t-1} + q$, we have $q_t = tq + q_0 = tq + 1 = tn_\xi + 1$ and:

$$x_t \sim csn_{p,q_t}(\mu_{x_t}, \Sigma_{x_t}, D_{x_t}, \vartheta_{x_t}, \Delta_{x_t})$$

3.1.2 State distribution - reduced rank case

In this paragraph I assume $G = [A|B]$ is full (row) rank (A can in principle be rank deficient). If it is not, before employing presented formulae, regularization described in section (2.3) needs to be applied first.

Assume that $x_{t-1} \sim csn_{p,q_{t-1}}(\mu_{x_{t-1}}, \Sigma_{x_{t-1}}, \Delta_{x_{t-1}}, \vartheta_{x_{t-1}}, \Delta_{x_{t-1}})$, which is true for $t = 1$, and $q = 1$, let $x_0 \sim csn_{p,1}(\mu_{x_0}, \Sigma_{x_0}, D_{x_0}, \vartheta_{x_0}, \Delta_{x_0})$ for[9]:

Since x_{t-1} and ξ_t, $t \in \mathcal{T}$, are independent variables, according to remark (2.16), joint distribution $g_t = (x_{t-1}, \xi_t)$ is:

$$g_t \sim csn_{p+n_\xi,q_t+q}(\mu_{g,t}, \Sigma_{g,t}, D_{g,t}, \vartheta_{g,t}, \Delta_{g,t})$$

[9]See the previous paragraph.

with parameters:

$$\mu_{g,t} = (\mu_{x_{t-1}}^T, \mu_\xi^T)^T, \quad \Sigma_{g,t} = \Sigma_{x_{t-1}} \oplus \Sigma_\xi, \quad D_{g,t} = D_{x_{t-1}} \oplus D_\xi,$$
$$\vartheta_{g,t} = (\vartheta_{x_{t-1}}^T, \vartheta_\xi^T)^T, \quad \Delta_{g,t} = \Delta_{x_{t-1}} \oplus \Delta_\xi$$

Under such notation, $x_t = Gg_t$ for $G = [A|B]$, and x_t follows:

$$x_t \sim csn_{p,q_t}(\mu_{x_t}, \Sigma_{x_t}, D_{x_t}, \vartheta_{x_t}, \Delta_{x_t})$$

for $q_t = q_{t-1} + q$, where, see remark (2.17):

$$\mu_{x,t} = G\mu_{g,t}, \quad \Sigma_{G,t} = G\Sigma_{g,t}G^T, \quad D_{x,t} = D_{g,t}\Sigma_{g,t}G^T\Sigma_{G,t}^{-1}$$
$$\vartheta_{x,t} = \vartheta_{g,t}, \quad \Delta_{x,t} = \Delta_{g,t} + D_{g,t}\Sigma_{g,t}D_{g,t}^T +$$
$$-D_{g,t}\Sigma_{g,t}G^T\Sigma_{G,t}^{-1}G\Sigma_{g,t}D_{g,t}^T$$

As in the previous paragraph, since $q_0 = 1$ and $q_t = q_{t-1} + q$, we have $q_t = tq + q_0 = tq + 1 = tn_\xi + 1$ and:

$$x_t \sim csn_{p,q_t}(\mu_{x_t}, \Sigma_{x_t}, D_{x_t}, \vartheta_{x_t}, \Delta_{x_t})$$

3.2 Observables

So far formulae for distribution of states x_t for all $t \in \mathcal{T}$ have been derived. Now let us do the distribution of observables y_t for all $t \in \mathcal{T}$. Once again, since normal distribution is a special case of *csn* distribution, it can be written that: $u_t \sim CSN_{n_u,1}(\mu_u, \Sigma_u, D_u, \vartheta_u, \Delta_u)$.

In DSGE, it usually is the case that $n_u = p$ and H is full rank, so that measurement errors rule out stochastic singularity in the measurement equation. If this is the case, which I assume, a matrix $K = [F, H]$ is full (row) rank, no matter what $r(F)$ is, and here no regularization must be involved.

Since $x_t \sim csn_{p,q_t}(\mu_{x_t}, \Sigma_{x_t}, \Delta_{x_t}, \vartheta_{x_t}, \Delta_{x_t})$[10] and x_t and ξ_t, $t \in \mathcal{T}$, are independent variables, according to remark (2.16), joint distribution $k_t = (x_t, u_t)$ is:

$$k_t \sim csn_{p+n_u,q_t+n_u}(\mu_{k,t}, \Sigma_{k,t}, D_{k,t}, \vartheta_{k,t}, \Delta_{k,t})$$

with parameters:

$$\mu_{k,t} = (\mu_{x_t}^T, \mu_u^T)^T, \quad \Sigma_{k,t} = \Sigma_{x_t} \oplus \Sigma_u, \quad D_{k,t} = D_{x_t} \oplus D_u,$$
$$\vartheta_{k,t} = (\vartheta_{x_t}^T, \vartheta_u^T)^T, \quad \Delta_{k,t} = \Delta_{x_t} \oplus \Delta_u$$

[10] See the previous paragraph.

Under such notation, $y_t = Kk_t$ for $K = [F|H]$, and y_t follows:

$$y_t \sim csn_{p,q_t+n_u}(\mu_{y_t}, \Sigma_{y_t}, D_{y_t}, \vartheta_{y_t}, \Delta_{y_t})$$

where, see remark (2.17):

$$\mu_{y,t} = K\mu_{k,t}, \quad \Sigma_{K,t} = K\Sigma_{k,t}G^T, \quad D_{y,t} = D_{k,t}\Sigma_{k,t}K^T\Sigma_{K,t}^{-1}$$
$$\vartheta_{y,t} = \vartheta_{k,t}, \quad \Delta_{y,t} = \Delta_{k,t} + D_{k,t}\Sigma_{k,t}D_{k,t}^T +$$
$$-D_{k,t}\Sigma_{k,t}K^T\Sigma_{K,t}^{-1}K\Sigma_{k,t}D_{k,t}^T$$

Notice that q-dimension of y_t is even bigger that of x_t - by the number of measurement errors $n_u = n$.

3.3 Conditional distributions

For $t \in \mathcal{T}$, let us define an information set $Y_t = \{y_1, y_2, ..., y_t\}$ which consists of observables up to time t. I will derive the *a posteriori* distribution $(x_t|Y_t)$ in a usual way, i.e. by constructing the joint distribution $(x_t, y_t|Y_{t-1})$ with the "residual trick" and than conditioning upon y_t.

Assume, that the *a posteriori* distribution of states x_{t-1}, i.e. conditional distribution of states x_{t-1} with respect to the information set Y_{t-1}, therefore after observing y_{t-1}, is given by:

$$(x_{t-1}|Y_{t-1}) \sim csn_{p,q_{t-1}}(\mu_{t-1}, \Sigma_{t-1}, D_{t-1}, \vartheta_{t-1}, \Delta_{t-1})$$

for some parameters μ_{t-1}, Σ_{t-1}, D_{t-1}, ϑ_{t-1} and Δ_{t-1}. If so, the *a priori* random variable $(x_t|Y_{t-1})$ is given by:

$$(x_t|Y_{t-1}) = (Ax_{t-1} + B\xi_t|Y_{t-1}) = (Ax_{t-1}|Y_{t-1}) + B\xi_t \sim csn(\mu, \Sigma, D, \vartheta, \Delta)$$

for:

$$\mu = \mu_A + \mu_B, \quad \Sigma = \Sigma_A + \Sigma_B, \quad D = \left[\Sigma_A D_A^T, \Sigma_B D_B^T\right]^T \Sigma^{-1}, \quad \vartheta = \left[\vartheta_A^T, 0^T\right]^T,$$
$$\Delta = \Delta_A \oplus \Delta_B + (D_A \oplus D_H)(\Sigma_A \oplus \Sigma_D)(D_A \oplus D_H)^T +$$
$$-(D_A\Sigma_A \oplus D_H\Sigma_H)(\Sigma)^{-1}(D_A\Sigma_A \oplus D_H\Sigma_H)^T$$

where:

$$\mu_A = A\mu_{t-1}, \quad \Sigma_A = A\Sigma_{t-1}A^T, \quad D_A = D_{t-1}\Sigma_{t-1}A^T\Sigma_A^{-1},$$
$$\vartheta_A = \vartheta_{t-1}, \quad \Delta = \Delta_{t-1} + D_{t-1}\Sigma_{t-1}D_{t-1}^T - D_{t-1}\Sigma_{t-1}A^T\Sigma_A^{-1}A\Sigma_{t-1}D_{t-1}^T$$

Observation equation in (1) defines a likelihood model for y_t, conditional on x_t and Y_{t-1}, which is is given by:

$$y_t = Fx_t + Hu_t$$

Then, according to corollary (2.18), the *a posteriori* distribution of $(x_t|Y_{t-1}$, conditional on y_t, i.e. distribution of $(x_t|Y_t)$ is as follows:

$$(x_t|Y_t) \sim csn_{n,q}(\mu_{x|y}, \Sigma_{x|y}, D_{x|y}, \vartheta_{x|y}, \Delta_{x|y})$$

where:

$$\mu_{x|y} = \mu + \Sigma F^T(F\Sigma F^T + H\Psi H^T)^{-1}(y - F\mu)$$

$$\Sigma_{x|y} = \Sigma - \Sigma F^T(F\Sigma F^T + H\Psi_u H^T)^{-1}H\Sigma$$

$$D_{x|y} = \left(\begin{pmatrix} D\Sigma \\ 0 \end{pmatrix} - \begin{pmatrix} D\Sigma F^T \\ 0 \end{pmatrix} (F\Sigma F^T + H\Psi_u H^T)^{-1}F\Sigma \right) \Sigma_{x|y}^{-1}$$

$$\vartheta_{x|y} = \begin{pmatrix} \vartheta \\ 0 \end{pmatrix} - \begin{pmatrix} D\Sigma F^T \\ 0 \end{pmatrix} (F\Sigma F^T + H\Psi_u H^T)^{-1}F\Sigma \ (y_t - F\mu)$$

$$\Delta_{x|y} = \begin{pmatrix} \Delta + D\Sigma D^T & 0 \\ 0 & I \end{pmatrix} - \begin{pmatrix} D\Sigma F^T \\ 0 \end{pmatrix} (F\Sigma F^T + H\Psi_u H^T)^{-1} \begin{pmatrix} D\Sigma F^T \\ 0 \end{pmatrix}^T - D_{x|y}\Sigma_{x|y}D_{x|y}^T$$

Now we can move on to the likelihood function, which, given formulae for distribution of variables $(x_t|Y_t)$, $t = 1, 2, ..., T$, is standard.

4 Likelihood function

The likelihood function of the state space model (1) is given by:

$$\mathcal{L} = p(y_0) \prod_{t=2}^{T} p(y_t|Y_{t-1})$$

Because Hu_t is independent of Y_{t-1}, using the measurement equation we get:

$$(y_t|Y_{t-1}) = (Fx_t + Hu_t|Y_{t-1}) = (Fx_t|Y_{t-1} + Hu_t) = F(x_t|Y_{t-1}) + H(u_t)$$

Since, in the notation of the previous paragraph, distribution of $(x_t|Y_{t-1})$ is:

$$(x_t|Y_{t-1}) \sim csn_{p,q_t}(\mu, \Sigma, D, \vartheta, \Delta)$$

the conditional *a priori* distribution of $(y_t|Y_{t-1}) = F(x_t|Y_{t-1}) + H(u_t)$ is:

$$(y_t|Y_{t-1}) \sim csn_{q_t+p}(\mu_y, \Sigma_y, D_y, \vartheta_y, \Delta_y)$$

with parameters:

$$\mu_y = \mu_F + \mu_H, \quad \Sigma_y = \Sigma_F + \Sigma_H, \quad D_y = \left[\Sigma_F D_F^T, \Sigma_H D_H^T\right]^T \Sigma_y^{-1},$$

$$\vartheta_y = \left[\vartheta_F^T, \vartheta_H^T\right]^T, \quad \Delta_y = \Delta_F \otimes \Delta_H + (D_F \otimes D_H)(\Sigma_F \otimes \Sigma_H)(D_F \otimes D_H)^T +$$

$$-\left(D_F\Sigma_F \oplus D_H\Sigma_H\right)(\Sigma_y)^{-1}\left(\Sigma_F D_F^T \oplus \Sigma_H D_H^T\right)$$

where:

$$\mu_F = F\mu_x, \quad \Sigma_F = F\Sigma_x F^T, \quad D_F = D_x\Sigma_x F^T\Sigma_F^{-1}$$

$$\vartheta_F = \vartheta_x, \quad \Delta_F = \Delta_x + D_x\Sigma_x D_x^T - D_x\Sigma_x F^T\Sigma_F^{-1}F\Sigma_x D_x^T$$

therefore:

$$p(y_t|Y_{t-1}) = \phi_p(y_t;\mu_y,\Sigma_y)\frac{\Phi_{q_t+p}(D_y(y_t-\mu_y);\vartheta_y,\Delta_y)}{\Phi_{q_t+p}(0;\vartheta_y,\Delta_y + D_y\Sigma_y D_y^T)} \tag{1}$$

Value of the likelihood function $\mathcal{L}_\theta = p(Y|\theta)$ can now be calculated for given θ. Value of $\mathcal{L}_\theta$ can be feeded into any numerical optimization routine. The model can be estimated.

5 Estimation example

As an exemplary empirical application of the estimation procedure we used a model described in Schorfheide (2000). It is a medium scale cash in advance DSGE economy with two drivers: the technology shock and the monetary policy shock (money stock growth rate shock). The technology shock $\epsilon_{A,t}$ enters the production function $Y_t = K_t^\alpha (A_t N_t)^{1-\alpha}$ in the following way:

$$\log A_t = \gamma + \log A_{t-1} + \epsilon_{A,t}$$

and the monetary policy shock $\epsilon_{M,t}$ disturbs the path of growth of money $m_t = \frac{M_{t+1}}{M_t}$ as follows:

$$\ln m_t = (1-\rho)\ln m^\star + \rho \ln m_{t-1} + \epsilon_{M,t}$$

where $m^\star$ denotes equilibrium growth rate of money supply in the economy. When normal distribution is used, we assume that:

$$\begin{pmatrix} \epsilon_{A,t} \\ \epsilon_{M,t} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_M^2 \end{pmatrix}\right)$$

parameter	distribution	mean	std. dev.
α	beta	0.356	0.02
β	beta	0.993	0.002
γ	normal	0.0085	0.003
$m^{\star}$	normal	1.0002	0.007
ρ	beta	0.129	0.223
ψ	beta	0.65	0.05
δ	beta	0.01	0.005
σ_A	inverteg gamma	0.035	0.0075
σ_M	inverteg gamma	0.009	inf
d_A	normal	0.00	60
d_M	normal	0.00	60

Table 1: Priors for the estimation.

whereas here we consider a more general setup:

$$\begin{pmatrix} \epsilon_{A,t} \\ \epsilon_{M,t} \end{pmatrix} \sim csn_{2,2}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_M^2 \end{pmatrix}, \begin{pmatrix} d_A & 0 \\ 0 & d_M \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

Note that the normal distribution case is also contained in the above specification (if $d_A = d_M = 0$).

The model was estimated using quarterly data for the US economy ranging from 1980Q1 to 2015Q1. Observable variables are real output growth rate $\Delta \ln GDP_t$ and inflation rate $\Delta \ln P_t$. Since there are two shocks and two observable variables, the likelihood function is non-degenerate. Altogether 12 parameters are estimated (the same which in Schorfheide (2000) plus shocks' skewness parameters d_A and d_M). Model was solved and estimated in Dynare, in which the Kalman filter routine was replaced by a skewed-Kalman filter and the space of parameters was extended so that it involved also skewness-related parameters d_A and d_M, which are also optimized along with the remaining 10 parameters.

A priori distributions of parameters were assumed as shown in Table (1). For skewness parameters d_A and d_M a relatively flat prior was assumed, with mean equal to zero and standard deviation equal to 60. This was the case, because we wanted

parameter	value	std. dev.
α	0.4099	0.0198
β	0.9939	0.0017
γ	0.0159	0.0011
$m^\star$	1.0125	0.0025
ρ	0.6978	0.0663
ψ	0.6068	0.0504
δ	0.0044	0.0025
σ_A	0.0167	0.0012
σ_M	0.0048	0.0003
d_A	-187.4056	30.3602
d_M	34.0910	51.8722

Table 2: Posterior modes and standard deviations.

the filter to identify skewness parameters without restrictions imposed on them *a priori*.

Table (2) shows estimation results obtained using the filter based on the closed skewed normal distribution.

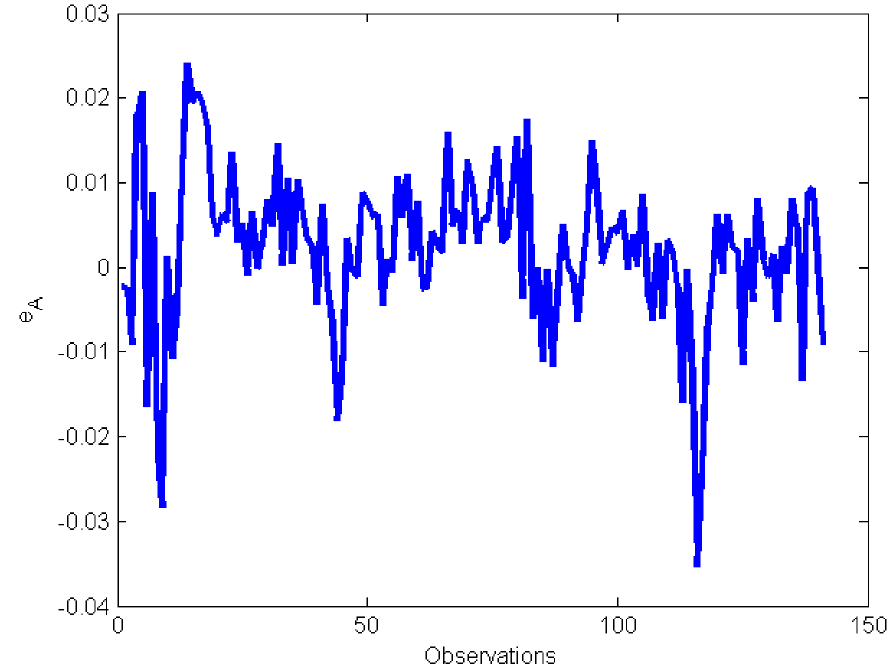

Figure 1: Time series of identified TFP shocks.

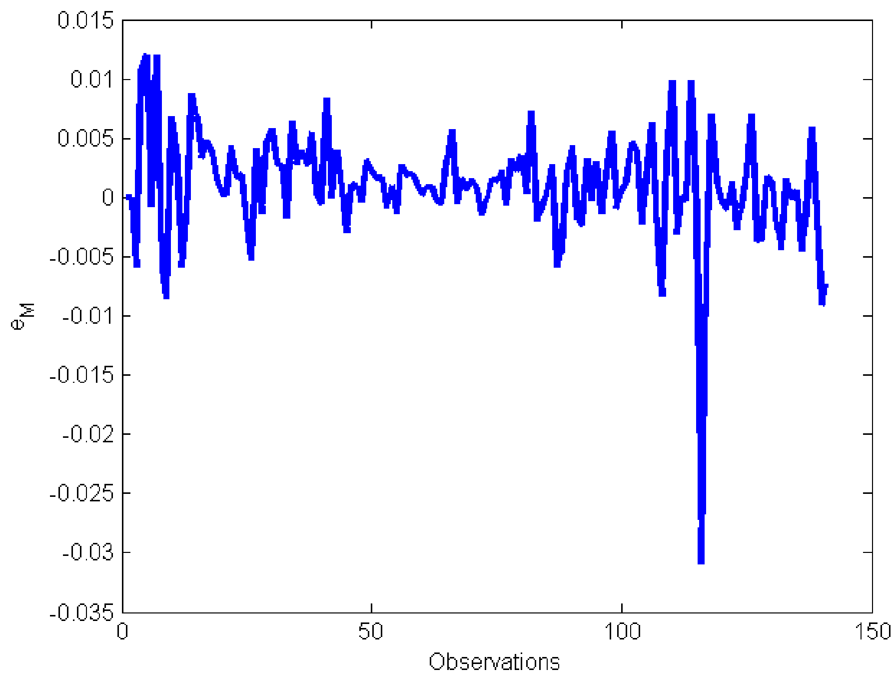

Figure 2: Time series of identified monetary shocks.

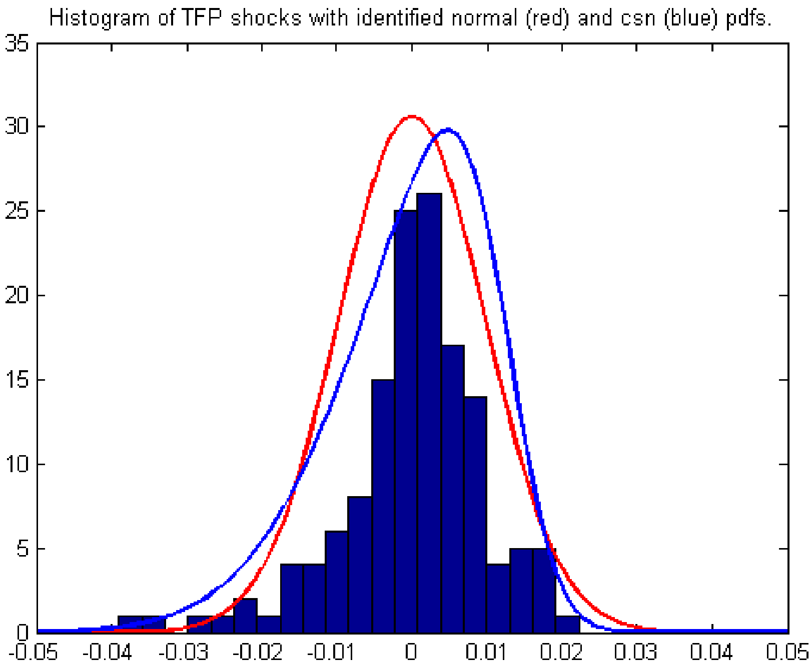

Figure 3: TFP shocks with a fitted normal pdf (red) and csn pdf (blue).

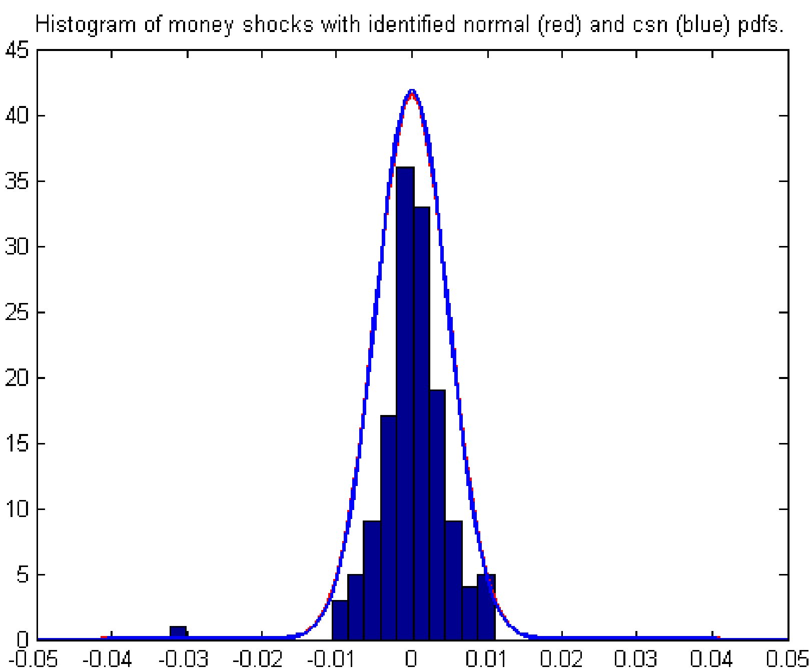

Figure 4: Money shocks with a fitted normal (red) and csn pdfs (blue).

Figures 1 and 2 present plots of identified shocks (ϵ_A and ϵ_M respectively), whereas Figures 3 and 4 presents histograms of these shocks with fitted normal distribution probability density functions and fitted closed skew-normal probability density functions[11]. While in case of TFP shocks the normal and the csn distributions differ substantially, for money shocks they practically overlap. This is in line with estimates of d_A and d_M. Skewness parameter of the TFP shock ($d_A = -187.4056$) is negative and significant. On the other hand, skewness parameter of the monetary shock ($d_M = 34.0910$) is positive, but not significant (ratio of the estimate to its standard deviation equals 0.66). Sample skewness of identified shocks ϵ_A equals -0.98 and of shocks ϵ_M is -1.95. Although sample skewness of ϵ_M is approximately two times bigger than that of ϵ_A, there is a reasons for which skewness of ϵ_M is not significant, and skewness of ϵ_A is. While skewness of ϵ_A is systematic, skewness of ϵ_M is attributed only to a single large negative outlier observation (see the histograms on Figure 3 and Figure 4). In fact, having removed this single observation, skewness of ϵ_M becomes positive (0.17). Forcing d_M to be negative, generates a large loss in the likelihood function by all the other observations.

Magnitude of d_A and d_M is bigger than that of the remaining parameters. This is due to the magnitude of standard deviations of shocks ϵ_M and ϵ_A, which is small, because sample skewness of a csn variable is implied by an interplay between the skewness parameter D and the covariance parameter Σ (which is not a covariance matrix of the distribution itself). This is most visible in the univariate case - the lower the variance parameter, the bigger the skewness parameter (up to a sign) is required to produce a given amount of skewness.

The model with skewness is favoured by the data. Laplace approximation of the log data density is 938.58, whereas for a model with normal shocks it equals 907.58. This should be the case, since the model with skewness has more degrees of freedom - it nests the normal one.

Were shocks distributed according to the normal distribution, i.e. with zero skewness, probability that a TFP shock falls below its 1st empirical percentile, equals about 2×10^{-4}, which means, that, on average, such a shocks should happen once every 1250 years (for data is quarterly). This observation is in line with Cúrdia,

[11] Pdfs were scaled so that they match with histograms on the figures.

Del Negro and Greenwald (2014), who point out, that the magnitude of identified shocks during recessions make such recessions, under normality assumption, almost impossible within a lifetime. As for the closed skew-normal distribution, analogical probability and frequency equal about 0.5% which translates into an event which happens once every 50 years, which seems to be a more plausible frequency.

References

[Azalini (1985)] Azzalini, A.: *A class of distributions which includes the normal ones*, Scandinavian Journal of Statistics (1985), Vol. 12, pp. 171–178.

[Azalini (1986)] Azzalini, A.: *Further results on a class of distributions which includes the normal ones*, Statistica (1986), Vol. 46, pp. 199–208.

[Azzalini and Capitanio (1999)] Azzalini, A. and A. Dalla Valle.: *Applications of the Multivariate Skew Normal Distribution*, Journal of the Royal Statistical Society. Series B (Statistical Methodology) (1999), Vol. 61, No. 3, pp. 579–602.

[Azalini and Dalla Valle (1996)] Azzalini, A. and A. Dalla Valle.: *The Multivariate Skew-Normal Distribution*, Biometrika (1996), Vol. 83, No. 4, pp. 715–726.

[Cúrdia, Del Negro and Greenwald (2014)] Cúrdia, V., Del Negro, M., Greenwald, D. L. (2014): *Rare Shocks, Great Recessions*, Journal of Applied Econometrics (2014), Vol. 29, No. 7, pp. 1032–1052.

[Genton (2001)] Genton, M. G., He, L. Liu, X., *Moments of skew-normal random vectors and their quadratic forms.*, Statistics & Probability Letters (2001), Vol. 51, pp. 319-325.

[Genton (2004)] Genton, M. G. (ed.): *Skew-elliptical distributions and their applications. A journel beyond normality.*, A CRC Press Company (2004).

[González-Farías et al. (2004)] González-Farías, G., Domínguez-Molina, J., Gupta, A. (2004): *Additive properties of skew normal random vectors.*, Journal of Statistical Planning and Inference, Vol. 126, pp. 521-534.

[Schorfheide (2000)] Schorfheide, F.: *Loss Function-Based Evaluation of DSGE Models*, Journal of Applied Econometrics, Vol. 15, pp. 645–670.

[Simon (2006)] Simon, D.: *Optimal State Estimation. Kalman, H_∞ and Nonlinear Approches*, John Willey & Sons (2006).

6 Proofs

Proof 6.1. *Proof of corollary (2.14).*

The mgf of Az is:

$$M_{Az}(t) = M_z(A^T t) = \frac{\Phi_q(D\Sigma A^T t; \vartheta, \Delta + D\Sigma D^T)}{\Phi_q(0; \vartheta, \Delta + D\Sigma D^T)} \mathrm{e}^{t^T A\mu + \frac{1}{2} t^T A\Sigma A^T t}$$

from which $\mu_A = A\mu$ and $\Sigma_A = A\Sigma A^T$. Note that since A is full (row) rank, then for $\Sigma > 0$ we have $\Sigma_A > 0$, so $D_A \Sigma_A t = D\Sigma A^T t$ can be expressed as:

$$D_A \Sigma_A t = D\Sigma A^T t = D\Sigma A^T \Sigma_A^{-1} \Sigma_A t$$

hence we can denote $D_A = D\Sigma A^T \Sigma_A^{-1}$. Also, $\Delta_A + D_A \Sigma_A D_A^T = \Delta + D\Sigma D^T$, hence:

$$\Delta_A = \Delta + D\Sigma D^T - D_A \Sigma_A D_A^T = \Delta + D\Sigma D^T - D\Sigma A^T (A\Sigma A^T)^{-1} A\Sigma^T D^T$$

Finally, we can denote $\vartheta_A = \vartheta$.

Proof 6.2. *Proof of corollary (2.18).*

The mgf of Az is:

$$M_{Az}(t) = M_z(A^T t) = \frac{\Phi_q(D\Sigma A^T t; \vartheta, \Delta + D\Sigma D^T)}{\Phi_q(0; \vartheta, \Delta + D\Sigma D^T)} \mathrm{e}^{t^T A\mu + \frac{1}{2} t^T A\Sigma A^T t}$$

from which $\mu_A = A\mu$ and $\Sigma_A = A\Sigma A^T$. Note that since A is full (column) rank, we have $A^T A > 0$, so $D_A \Sigma_A t = D_A A\Sigma A^T t = D\Sigma A^T t$ can be expressed as:

$$D_A A\Sigma A^T t = D(A^T A)^{-1} A^T A\Sigma A^T t$$

hence we can denote $D_A = D(A^T A)^{-1} A^T$. Also:

$$\Delta_A + D_A \Sigma_A D_A^T = \Delta_A + D(A^T A)^{-1} A^T A\Sigma A^T A (A^T A)^{-1} D^T = \Delta_A + D\Sigma D^T$$

hence we can denote $\Delta_A = \Delta$. Finally, we can denote $\vartheta_A = \vartheta$.

Proof 6.3. *Proof of corollary (2.16).*

Since z_i, $i = 1, 2, \ldots, n$ are independent, probability density function of $z = (z_1^T, z_2^T, \ldots, z_n^T)^T$

is given by:

$$z \sim \prod_{i=1}^{n} \phi_{p_i}(z_i;\mu_i,\Sigma_i)\frac{\Phi_{q_i}(D_i(z_i-\mu_i);\vartheta_i,\Delta_i)}{\Phi_{q_i}(0;\vartheta_i,\Delta_i+D_i\Sigma_i D_i^T)} =$$

$$\prod_{i=1}^{n} \phi_{p_i}(z_i;\mu_i,\Sigma_i)\frac{\prod_{i=1}^{n}\Phi_{q_i}(D_i(z_i-\mu_i);\vartheta_i,\Delta_i)}{\prod_{i=1}^{n}\Phi_{q_i}(0;\vartheta_i,\Delta_i+D_i\Sigma_i D_i^T)} =$$

$$\phi_{\sum_{i=1}^{n} p_i}(z;\mu^\star,\Sigma^\star)\frac{\Phi_{\sum_{i=1}^{n} q_i}(D^\star(z-\mu^\star);\vartheta^\star,\Delta^\star)}{\Phi_{\sum_{i=1}^{n} q_i}(0;\vartheta^\star,\Delta^\star+D^\star\Sigma^\star(D^\star)^T)}$$

where:

$$\mu^\star = (\mu_1^T,\mu_2^T,...,\mu_n^T)^T, \quad \Sigma^\star = \oplus_{i=1}^{n}\Sigma_i, \quad D^\star = \oplus_{i=1}^{n} D_i,$$

$$\vartheta^\star = (\vartheta_1^T,\vartheta_2^T,...,\vartheta_n^T)^T, \quad \Delta^\star = \oplus_{i=1}^{n}\Delta_i$$

Proof 6.4. *Proof of corollary (2.17).*

Proof of corollary (2.17) follows directly from corollaries (2.14) and (2.16), since for $z = (z_1^T, z_2^T, \ldots, z_n^T)^T$ *and* $A = 1_n^T \otimes I_p$ *we have* $\sum_{i=1}^{n} z_i = Az$*, where* $A \in \mathrm{R}^{p\times np}$ *and* $r(A) = p$*, so that* A *is full (row) rank.*