Maximum likelihood estimation of state space models with skewed shocks

Grzegorz Koloch∗†

July 23, 2012

Abstract

This paper provides formulae for the likelihood function, filtration densities and prediction densities of linear state space models in which shocks are allowed to be skewed. In particular I work with the closed skew normal distribution introduced in Genton (2004), which nests a normal distribution as a special case. Closure of the distribution with respect to all necessary transformations in the state space setting is guaranteed by a simple regularization 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 observed data, in shocks, impulse responses and forecasts confidence intervals. Some of the advantages of using the outlined approach involve capturing asymmetric inflation risks in central banks forecasts or producing more plausible probabilities of deep but rare recessionary episodes with DSGE/VAR filtration.

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

JEL: C51, C13, E32

∗National Bank of Poland, Economic Institute, Bureau of Applied Research, Swietokrzyska 19/21, 00-919 Warsaw, Poland.
†Warsaw School of Economics, Institute of Econometrics, Department of Decision Analysis and Support, Al. Niepodleglosci 164, 00-950 Warsaw, Poland.
1 Introduction

In this paper I provide a maximum likelihood estimator for a linear state space model with skewed shocks in the transition equation\(^1\). Let us consider the following model:

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

for \(t \in T = \{1, 2, ..., T\}\), where \(x_t \in \mathbb{R}^p\) and \(y_t \in \mathbb{R}^n\) denote states and observables respectively, \(\xi_t \in \mathbb{R}^{n_\xi}\) and \(u_t \in \mathbb{R}^{n_u}\), \(n_\xi, n_u \geq 1\), denote shocks and measurement errors respectively, \(R^{p \times p} \ni A \neq 0, B \in R^{p \times n_\xi}, F \in R^{n \times p}, H \in R^{p \times n_u}\), moreover \(\Psi_u \in R^{n_u \times n_u}, \Psi_{x_0} \in R^{p \times p}\), and \(|\Psi_u|, |\Psi_{x_0}| \geq 0\). Finally, \(p_\xi(\theta_\xi)\) denotes a probability distribution function of martingale difference shocks \(\xi_t\), which depends on a vector of parameters \(\theta_\xi\).

A usual assumption is that \(p_\xi\) is a multivariate normal distribution, independent across its dimensions. In such a case the Kalman filter constitutes an optimal filter\(^2\), see Simon (2006). If normality assumption is relaxed, Kalman filter remains an optimal linear filter. In this article, I assume that elements of \(\xi_t\) are independent, but, for some values of \(\theta_\xi\), \(p_\xi\) represents a skewed (asymmetric) distribution. In particular, I assume that shocks \(\xi_t\) follow a closed skew normal distribution (csn henceforth), which nests the normal distribution as a special case, see Genton et al. (2004). The csn distribution is chosen, because it is closed under most 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 deficiency of matrix \(A\)\(^4\). This case turns out to be an obstacle in ML estimation since singularity of \(A\) precludes

\(^1\)Measurement shocks are assumed to be normally distributed but extension to skewed measurement errors is straightforward. I prefer to work with normal measurement errors because of the appealing interpretation based on the central limit theorem.

\(^2\)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.

\(^4\)Which is a standard case when the state space form represents a first order perturbation of a DSGE model.
propagation of the *csn* distribution through the state space setting. I use a simple
regularization procedure to deal with this problem.

To allow for prediction, filtration and estimation I 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 $\theta = (\theta_F, \theta_H, \theta_u, \theta_A, \theta_F, \theta_x, \theta_o)$ and $Y_t = \{y_t, y_{t-1}, \ldots, y_1\}$. When one goes through the article, one should also find it easy to obtain formulae for any type of filtration, eg. for smoothing. Although the formulae are provided, some of them involve potentially vary large scale normal integrals, practical calculation of which is computationally demanding. To make them operational I factor multivariate integrals into the products of univariate ones. Consequences of this approximation for estimation are left for further research.

Since state space form corresponds to some popular macroeconomic tools – VARs and reduced form first order perturbations of DSGE models, results contained in the paper allow for extending the agenda of empirical macroeconomics by issues related to skewness in the data in an computationally efficient manner, ie. without the need of resorting to sampling techniques. In particular, estimation of VARs and DSGE models with skewed shocks becomes analagical to Kalman filter ML estimation, while skewness of observables, states, confidence intervals of impulse responses and of forecasts can be assessed without referring to nonlinearities in the model. Some of the advantages of using the outlined approach involve capturing asymmetric inflation risks in central banks forecasts or producing more plausible probabilities of deep but rare recessionary episodes with VAR or DSGE. In this article I focus on econometric part of the agenda while leaving empirical applications for further research.

Remainder of the paper is arranged as follows. In section 2 I describe the closed skewed normal distribution introduced in Genton et al. (2004) and discuss its basic properties. Section 3 provides the filter and section 4 the likelihood function.

# 2 The skewed normal distribution

## 2.1 Definition

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

\(^5\)All vectors are column vectors in this paper.
cumulative distribution function of a \( q \)-dimensional normal distribution with mean \( \mu \) and nonnegative-definite covariance matrix \( \Sigma \) by \( \Phi_q(z; \mu, \Sigma) \). For \( q > 1 \) function \( \Phi_q \) does not have a closed form. Let \( r(M) \) denote a rank of \( M \). I will define the closed skewed normal, possibly singular, distribution by means of the moment generating function (mgf) and then, under nonsingularity of covariance, by the probability density function (pdf).

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

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

which henceforth will be denoted by:

\[
z \sim \text{csn}_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)
\]

If \( |\Sigma| > 0 \), a \( \text{csn} \) 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 \( \mu, \Sigma, D, \vartheta \) and \( \Delta \), where \( \mu \in \mathbb{R}^p \), \( \vartheta \in \mathbb{R}^q \), \( \Sigma \in \mathbb{R}^{p \times p} \), \( |\Sigma| > 0 \), \( \Delta \in \mathbb{R}^{q \times q} \), \( |\Delta| \geq 0 \) and \( D \in \mathbb{R}^{q \times p} \), then 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 (2) 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 \( \mu, \Sigma, D, \vartheta \) and \( \Delta \) if and only if its density function for every \( z \in \mathbb{R}^p \) equals \( p(z) \) in (2). Parameters \( \mu, \Sigma \) and \( D \) have interpretation of location, scale and skewness parameters respectively. Parameters \( \vartheta \) and \( \Delta \) are artificial, but inclusion of them allows for closure of the \( \text{csn} \) distribution under conditioning and marginalization respectively. The \( q \)-dimension in \( \Phi_q \) is also artificial, but it allows for closure of sums and for taking the joint distribution of independent (not necessarily \( iid \)) variables. When \( \Sigma, D \) and \( \Delta \) are scalars,
they will be denoted respectively by $\sigma$, $d$ and $\delta$. For $D = 0$, the $csn$ distribution reduces to a $p$-dimensional normal one.

### 2.2 Properties

This section discusses properties of the $csn$ distribution which I will need in the remainder of the paper. I will concentrate on two critical issues - the closure of the distribution and large scale normal integration (the $q$-dimension expansion). First, however, all relevant remarks and corollaries are outlined.

#### 2.2.1 Remarks, corollaries.

**Remark 2.3.** For $p = q = 1$, $\vartheta = 0$ and $\Delta = 1$ the $csn$ distribution reduces to the Azzalini skewed normal distribution, see Azzalini (1996), Azzalini (1999).

Such a case will be denoted by:

$$z \sim sn(\mu, \sigma, d)\quad (3)$$

**Corollary 2.4.** Let $z \sim csn_{1,1}(\mu, \sigma, d, \vartheta, \delta)$ for parameters as in definition (2.2), and assume that $\delta + d^2\sigma \neq 0$, then:

$$E(z) = \mu + \sqrt{\frac{2}{\pi}} \frac{d\sigma}{\sqrt{\delta + d^2\sigma}}\quad (4)$$

$$D(z) = \sigma - \frac{2}{\pi} \frac{d^2\sigma^2}{\delta + d^2\sigma}\quad (5)$$

$$E(z - E(z))^3 = \left(2 - \frac{\pi}{2}\right) \left(\sqrt{\frac{2}{\pi}}\right)^3 \left(\frac{d\sigma}{\delta + \sigma d^2\sigma}\right)^3\quad (6)$$

It follows that:

**Remark 2.5.** Let $z \sim csn_{1,1}(\mu, \sigma, d, \vartheta, \delta)$ for parameters as in corollary (2.4), then $E(z) = 0$ if and only if $\mu = -\sqrt{\frac{2}{\pi}} \frac{d\sigma}{\sqrt{\delta + d^2\sigma}}$.

**Corollary 2.6.** Let $z \sim csn_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)$, $p, q \geq 1$, for parameters as in definition (2.2.2). Elements of $z$ are independent if and only if matrices $\Sigma$ and $D$ are diagonal.

**Corollary 2.7.** Let $z \sim csn_{p,q}(\mu, \Sigma, D, \vartheta, \Delta)$, $p, q \geq 1$, for parameters as in definition (2.2.2). 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 2.8. Let \( z \sim \text{csn}_{1,q}(\mu, \sigma, d, \vartheta, \delta) \), \( q \geq 1 \) and for parameters as in definition (2.2.2), let also \( p \neq 0 \) and \( b \in \mathbb{R} \), then:

\[
\rho z + b \sim \text{csn}_{1,q}(\rho \mu + b, \rho^2 \sigma, \frac{1}{\rho} d, \vartheta, \delta)
\]

Corollary 2.9. Let \( z \sim \text{csn}_{p,q}(\mu, \Sigma, D, \vartheta, \Delta) \), \( p, q \geq 1 \), for parameters as in definition (2.2.2), let also \( A \in \mathbb{R}^{p \times p} \), \( A \neq 0 \), and \( b \in \mathbb{R}^p \), then:

\[
Az + b \sim \text{csn}_{p,q}(A\mu + b, A\Sigma A^T, D\Sigma A^{-1}, \vartheta, \Delta)
\]

Corollary 2.10. Let \( z_i \sim \text{csn}_{p_i,q_i}(\mu_i, \Sigma_i, D_i, \vartheta_i, \Delta_i) \), \( p_i, q_i \geq 1 \), \( i = 1, 2, ..., n \), for parameters as in definition (2.2.2), then \( \sum_{i=1}^n z_i \sim \text{csn}_{p,\Sigma_{i=1}^n q_i}(\mu^*, \Sigma^*, D^*, \vartheta^*, \Delta^*) \), where:

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

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

for \( \Delta^{\oplus} = \bigoplus_{i=1}^n \Delta_i \), \( D^{\oplus} = \bigoplus_{i=1}^n D_i \) and \( \Sigma^{\oplus} = \bigoplus \Sigma_i \), where operator \( \oplus \), for arbitrary matrices \( A \) and \( B \), is defined as:

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

Corollary 2.11. Let \( z_i \sim \text{csn}_{p_i,q_i}(\mu_i, \Sigma_i, D_i, \vartheta_i, \Delta_i) \), \( p_i, q_i \geq 1 \), \( i = 1, 2, ..., n \), for parameters as in definition (2.2.2), then the joint random variable: \( (z_1^T, z_2^T, ..., z_n^T)^T \sim \text{csn}_{p,\Sigma_{i=1}^n q_i}(\mu^*, \Sigma^*, D^*, \vartheta^*, \Delta^*) \), where:

\[
\mu^* = (\mu_1^T, \mu_2^T, ..., \mu_n^T)^T, \quad \Sigma^* = \bigoplus_{i=1}^n \Sigma_i, \quad D^* = \bigoplus_{i=1}^n D_i,
\]

\[
\vartheta^* = (\vartheta_1^T, \vartheta_2^T, ..., \vartheta_n^T)^T, \quad \Delta^* = \bigoplus_{i=1}^n \Delta_i
\]

2.2.2 Closure and the regularization procedure

Note that in definition matrix \( \Sigma \) is allowed to be singular. If \( \Sigma \) is not positive definite, i.e. \( |\Sigma| = 0 \), resulting distribution is called singular. If \( \Sigma \) 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. Under full rank transformation I mean full row 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.
full rank) linear transformations transform nonsingular \( csn \) variables into nonsingular ones and singular variables into singular ones. Full row, but column rank deficient linear transformations (eg. dimension shrinkage/reduction) transform nonsingular \( csn \) variables into nonsingular ones and singular variables into singular or nonsingular 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 if arbitrary linear transformations are allowed. This fact is negative for ML estimation when the transition matrix \( A \) in state space equations is singular, which typically is the case in DSGE modeling.

In what follows I discuss two remarks in this respect. First, I 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 regularization procedure is suggested.

**Proposition 2.1.** Let \( \eta \in \mathbb{R}^m \) be distributed according to a \( csn_{m,q} \) for some \( p,q \geq 1 \) with parameters \( \mu_\eta, \Sigma_\eta \geq 0, D_\eta, \vartheta_\eta \) and \( \Delta_\eta > 0 \). Let \( z = G\eta, G \in \mathbb{R}^{p \times m} \). Then, \( z_t \) has a \( csn \) (possibly singular) distribution if and only if \( G \) has full row rank or if \( r(G^T) = r(G^T[w_i]) \) for all \( i = 1, 2, ..., q \), where \( r(G) \) denotes rank of \( G \) and \( w_i \) denotes the \( i \)-th row \( D_\eta \).

Proposition (2.1) states, that for a \( csn \) variable \( \eta \), 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 \( D_\eta \) be linear combinations of rows of \( G \), in other words, that rows of \( D_\eta \) 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 \( D_\eta \) – 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 \( D_\eta \) can be in principle arbitrary.

The correspondence between proposition (2.1) and the state space formulation
\[ z = x_t \]
\[ G = [A|B] \]
\[ \eta = (x_{t-1}^T, \xi_t^T)^T \]

Since \( p < m \), according to proposition (2.1), for the \textit{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 obviously does not affect the value of the likelihood function and, moreover, this can be done automatically.

Let us denote by \( \tilde{x}_t \) the (any) maximal linearly independent subset of states from \( x_t \), and by \( \check{x}_t \) the remaining states. Numerically, we can find a matrix \( K \) such that \( \check{x}_t = K\tilde{x}_t \) (for all \( t \in T \)). Rearrange \( x_t = [\tilde{x}_t^T, \check{x}_t^T] \) and partition model matrices accordingly. Then, first two state space equations in (1) can be rewritten as follows:

\[ y_t = \bar{F}\tilde{x}_t + Hu_t \]
\[ \tilde{x}_t = \bar{A}\tilde{x}_{t-1} + \bar{B} \]

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 \( \check{x}_t = K\tilde{x}_t \), we also have full information about the states which were substituted for.

\[ 2.2.3 \quad \text{q-dimension expansion} \]

Because \( \Sigma \) is a \( p \times p \) matrix and \( D \) is a \( q \times p \) matrix, corollary (2.6) 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\(^7\) \( \xi_t \). Corollary 2.10 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 \( \xi_t \), therefore, according to corollary (2.6), contribution of \( \xi_t \) to the \( q \)-dimension of \( x_t \) in every period cannot be squeezed to eg. 1, but must be equal to the size of \( \xi_t \) (i.e. \( n_{\xi} \)), hence the \( q \)-dimension of \( x_t \) quickly increases with \( t \). Let us then note, that moments of the multivariate \( csn \) distribution, probability density function (2) and, as a consequence, the likelihood function provided in section 4, they all involve a cumulative probability distribution function of a \( q \)-dimensional normal distribution.

To my best knowledge there is no efficient way of calculating large scale normal integrals with an arbitrary correlation structure, therefore I work with the following approximation \( \Phi_q(z) \approx \prod_{j=1}^q \Phi_1(z_j) \). Relevance of such an approximation in the context of issues related to skewness identification is left for further research.

3 The filter

In this section I 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 \( \theta = (\theta_F, \theta_H, \theta_u, \theta_A, \theta_F, \theta\xi, \theta_0) \). First, however, I derive unconditional distributions for states and observables.

The state space setting is assumed to be (1) and distribution \( p_\xi \) of shocks \( \xi_t \), \( t \in T \), is assumed to be a multivariate independent \( csn \) distribution:

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

for \( t = 1, 2, ..., T \) with \( \xi_t(i) \sim sd(\mu_\xi_i, \sigma_\xi_i, d_\xi_i) \). Remark (2.6) implies that matrices \( \Sigma_\xi \) and \( D_\xi \) are diagonal. For \( D_\xi \) 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 by \( n_{\xi} \), since to have \( E(\xi_t) = 0 \), one needs to impose: \( \mu_\xi_i = -\sqrt{\frac{2}{\pi}} \frac{d_\xi_i \sigma_\xi_i}{\sqrt{d_\xi_i^2 + \sigma_\xi_i^2}} \), \( i = 1, 2, ..., n_{\xi} \).

\(^7\)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).
where $\mu_{\xi}$ is the $i$-th element of $\mu_{\xi}$, $\sigma_{\xi}$ is the $i$-the diagonal element of $\Sigma_{\xi}$ and $\delta_{\xi} = 1$ is the $i$-th diagonal element of $\Delta_{\xi}$. 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 I 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. I 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 I assume $A$ is full rank. If $n_{\xi} < p$, then $B_{\xi t}$ does not have a distribution function since matrix $B_{\xi} \Sigma_{\xi} B_{\xi}^T$ 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:

$$\mu_{B} = B_{\mu_{\xi}}, \quad \Sigma_{B} = B_{\Sigma_{\xi}} B_{\xi}^T, \quad D_{B} = D_{\xi}(B_{\xi}^T B_{\xi})^{-1} B_{\xi}^T, \quad \vartheta_{B} = \vartheta_{\xi}, \quad \Delta_{B} = \Delta_{\xi}$$

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:

$$x_0 \sim csn_{p,1}(\mu_{x_0}, \Sigma_{x_0}, D_{x_0}, \vartheta_{x_0}, \Delta_{x_0})$$
Generally, i.e. for \( t \in T \), if:

\[
x_{t-1} \sim \text{csn}_{p,q_{t-1}}(\mu_{x_{t-1}}, \Sigma_{x_{t-1}}, \vartheta_{x_{t-1}}, \Delta_{x_{t-1}})
\]

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

\[
Ax_{t-1} \sim \text{csn}_{p,q_{t-1}}(\mu_{A,t-1}, \Sigma_{A,t-1}, \vartheta_{A,t-1}, \Delta_{A,t-1})
\]

where:

\[
\begin{align*}
\mu_{A,t-1} &= A \mu_{x,t-1}, \\
\Sigma_{A,t-1} &= A \Sigma_{x,t-1} A^T, \\
D_{A,t-1} &= D_{x,t-1} \Sigma_{x,t-1} A^T \Sigma_{A,t-1}^{-1}, \\
\vartheta_{A,t-1} &= \vartheta_{x,t-1}, \\
\Delta_{A,t-1} &= \Delta_{x,t-1} + D_{x,t-1} \Sigma_{x,t-1} D_{x,t-1}^T + \\
&- D_{x,t-1} \Sigma_{x,t-1} A^T \Sigma_{A,t-1}^{-1} A \Sigma_{x,t-1} D_{x,t-1}^T
\end{align*}
\]

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

\[
x_t \sim \text{csn}_{p,q_t}(\mu_{x_t}, \Sigma_{x_t}, D_{x_t}, \vartheta_{x_t}, \Delta_{x_t})
\]

for \( q_t = q_{t-1} + q \), where, see remark (2.10):

\[
\begin{align*}
\mu_{x,t} &= \mu_{A,t} + \mu_B, \\
\Sigma_{x,t} &= \Sigma_{A,t} + \Sigma_B, \\
D_{x,t} &= \left[ \Sigma_{A,t} D_{A,t}^T, \Sigma_B D_B^T \right] T \Sigma_{x,t}^{-1}, \\
\vartheta_{A,t} &= \left[ \vartheta_{A,t}^T, \vartheta_{B,t}^T \right]^T, \\
\Delta_{x,t} &= \Delta_{A,t} \otimes \Delta_B + (D_{A,t} \otimes D_B)(\Sigma_{A,t} \otimes \Sigma_B)(D_{A,t} \otimes D_B)^T + \\
&- (D_{A,t} \otimes D_B) \Sigma_{x,t}^{-1} (\Sigma_{A,t} D_{A,t}^T \otimes \Sigma_B D_B^T)
\end{align*}
\]

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

\[
x_t \sim \text{csn}_{p,q_t}(\mu_{x_t}, \Sigma_{x_t}, D_{x_t}, \vartheta_{x_t}, \Delta_{x_t})
\]

\(^8\Delta_{x_0} > 0 \) can in fact be arbitrary.
### 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 $x_{t-1} \sim \text{csn}_{p,q_{t-1}}(\mu_{x_{t-1}}, \Sigma_{x_{t-1}}, \Delta_{x_{t-1}}, \vartheta_{x_{t-1}}, \Delta_{x_{t-1}})$, which is true for $t = 1$, and $q = 1$, let $x_0 \sim \text{csn}_{p,1}(\mu_{x_0}, \Sigma_{x_0}, D_{x_0}, \vartheta_{x_0}, \Delta_{x_0})$ for $9$.

Since $x_{t-1}$ and $\xi_t$, $t \in T$, are independent variables, according to remark (2.11), joint distribution $g_t = (x_{t-1}, \xi_t)$ is:

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

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, $x_t = G g_t$ for $G = [A|B]$, and $x_t$ follows:

$$x_t \sim \text{csn}_{p,q}(\mu_{x_t}, \Sigma_{x_t}, D_{x_t}, \vartheta_{x_t}, \Delta_{x_t})$$

for $q_t = q_{t-1} + q$, where, see remark (2.10):

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

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

$$-D_{g,t} \Sigma_{g,t} G^T \Sigma_{G,t}^{-1} G \Sigma_{g,t} D_{g,t}^T$$

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:

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

### 3.2 Observables

So far formulae for distribution of states $x_t$ for all $t \in T$ have been derived. Now let us do the distribution of observables $y_t$ for all $t \in T$. Once again, since normal distribution is a special case of $\text{csn}$ distribution, it can be written that:

$$u_t \sim \text{CSN}_{n_u,1}(\mu_u, \Sigma_u, D_u, \vartheta_u, \Delta_u)$$

---

9See the previous paragraph.
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 \( x_t \sim \text{csn}_{p,q}(\mu_{x_t}, \Sigma_{x_t}, \Delta_{x_t}, \vartheta_{x_t}, \Delta_{x_t}) \) and \( x_t \) and \( \xi_t \) are independent variables, according to remark (2.11), joint distribution \( g_t = (x_t, u_t) \) is:

\[
k_t \sim \text{csn}_{p+nu+nu}(\mu_{k,t}, \Sigma_{k,t}, \Delta_{k,t}, \vartheta_{k,t}, \Delta_{k,t})
\]

with parameters:

\[
\mu_{k,t} = (\mu_{x,t}^T, \mu_{u,t}^T)^T, \quad \Sigma_{k,t} = \Sigma_{x_t} + \Sigma_{u}, \quad D_{k,t} = D_{x_t} + D_{u}, \quad \vartheta_{k,t} = (\vartheta_{x,t}^T, \vartheta_{u,t}^T)^T, \quad \Delta_{k,t} = \Delta_{x_t} + \Delta_{u}
\]

Under such notation, \( y_t = K k_t \) for \( K = [F|H] \), and \( y_t \) follows:

\[
y_t \sim \text{csn}_{p,q}(\mu_{y,t}, \Sigma_{y,t}, D_{y,t}, \vartheta_{y,t}, \Delta_{y,t})
\]

where, see remark (2.10):

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

Notice that \( q \)-dimension of \( y_t \) is even bigger that of \( x_t \) - by the number of observed variables \( n_u = n \).

### 3.3 Conditional distributions

For \( t \in T \), 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 \textit{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 \textit{a posteriori} distribution of states \( x_{t-1} \), i.e. conditional distribution of states \( x_t \) with respect to the information set \( Y_{t-1} \), therefore after observing \( y_{t-1} \), is given by:

\[
(x_{t-1}|Y_{t-1}) \sim \text{csn}_{p,q}(\mu_{t-1}, \Sigma_{t-1}, D_{t-1}, \vartheta_{t-1}, \Delta_{t-1})
\]

\(^{10}\)See the previous paragraph.
for some parameters $\mu_{t-1}$, $\Sigma_{t-1}$, $D_{t-1}$, $\vartheta_{t-1}$ and $\Delta_{t-1}$. If so, the \textit{a priori} random variable $(x_t|Y_{t-1})$ is given by:

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

for:

$$
\mu = \mu_A + \mu_B,

\Sigma = \Sigma_A + \Sigma_B,

D = [\Sigma_AD_A^T, \Sigma_BD_B^T]^T \Sigma^{-1},

\vartheta = [\vartheta_A^T, 0^T]^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},

\Sigma_A = A\Sigma_{t-1}A^T,

D_A = D_{t-1}\Sigma_{t-1}A^T\Sigma_A^{-1},

\vartheta_A = \vartheta_{t-1},

\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
$$

Information contained in the joint distribution of $(x_t, y_t|Y_{t-1})$ is equivalent to the information contained in the distribution of $(x_t, e_t|Y_{t-1})$ where:

$$
e_t = y_t - E(y_t|Y_{t-1})
$$

which is approximated by\textsuperscript{11}:

$$
e_t = y_t - F\mu
$$

Distribution of $e_t$ conditional on $x_t$ and $Y_{t-1}$ is normal:

$$
(e_t|x_t, Y_{t-1}) = (Fx_t + Hu_t - F\mu|x_t, Y_{t-1}) = (F(x_t - \mu) + (Hu_t|Y_{t-1})) =

= (F(x_t - \mu) + Hu_t) = F(x_t - \mu) + H(u_t) \sim N(F(x_t - \mu), \Sigma_H)
$$

Therefore, joint distribution of $(x_t, e_t|Y_{t-1})$, which contains the same information as distribution of $(x_t, y_t|Y_{t-1})$ is given by:

$$(x_t, e_t|Y_{t-1}) \sim \text{csn}_{p+n,q_t+q_u+n}(\mu_{(x_t,e_t)}, \Sigma_{(x_t,e_t)}, D_{(x_t,e_t)}, \vartheta_{(x_t,e_t)}, \Delta_{(x_t,e_t)})$$

where:

$$
\mu_{(x_t,e_t)} = \begin{bmatrix} \mu \\ 0 \end{bmatrix},

\Sigma_{(x_t,e_t)} = \begin{bmatrix} \Sigma & \Sigma F^T \\ F^T\Sigma & F\Sigma F^T + \Sigma_H \end{bmatrix},

D_{(x_t,e_t)} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix},

\vartheta_{(x_t,e_t)} = \begin{bmatrix} \vartheta \\ 0 \end{bmatrix},

\Delta_{(x_t,e_t)} = \Delta
$$

\textsuperscript{11}Approximation is only for simplicity, true expected value can be derived explicitly from state-space equations.
Therefore, the \textit{a posteriori} distribution, i.e. distribution of \((x_t|Y_t)\) is:

\[(x_t|Y_t) \sim csn(\mu_t, \Sigma_t, D_t, \vartheta_t, \Delta_t)\]

with parameters:

\[
\begin{align*}
\mu_t &= \mu + \Sigma F^T (F^T \Sigma F + \Sigma_H)^{-1} (y_t - F \mu),
\Sigma_t &= \Sigma - \Sigma F^T (F^T \Sigma F + \Sigma_H)^{-1} F \Sigma T, \\
D_t &= D, \quad \vartheta_t = \vartheta, \quad \Delta_t = \Delta
\end{align*}
\]

Now I 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 \textit{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:

\[
\begin{align*}
\mu_y &= \mu_F + \mu_H, \quad \Sigma_y = \Sigma_F + \Sigma_H, \\
D_y &= [\Sigma_F D_F^T, \Sigma_H D_H^T]^T \Sigma_y^{-1}, \\
\vartheta_y &= [\vartheta_F^T, \vartheta_H^T]^T, \quad \Delta_y = \Delta_F \otimes \Delta_H + (D_F \otimes D_H)(\Sigma_F \otimes \Sigma_H)(D_F \otimes D_H)^T + \\
&\quad (D_F \Sigma_F \otimes D_H \Sigma_H) (\Sigma_y)^{-1} (\Sigma_F D_F^T \otimes \Sigma_H D_H^T)
\end{align*}
\]

where:

\[
\begin{align*}
\mu_F &= F \mu_x, \quad \Sigma_F = F \Sigma_x F^T, \\
D_F &= D_x \Sigma_x D_x^T F \Sigma_F^{-1} \\
\vartheta_F &= \vartheta_x, \quad \Delta_F = \Delta_x + D_x \Sigma_x D_x^T - D_x \Sigma_x F^T \Sigma_F^{-1} F \Sigma_x D_x^T
\end{align*}
\]
therefore:

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

(8)

Value of the likelihood function $L_\theta = p(Y | \theta)$ can now be calculated for given $\theta$. Value of $L_\theta$ can be fed into any numerical optimization routine. The model can be estimated.

References


