APPENDIX: A Class of Non-Gaussian State Space Models with Exact Likelihood Inference

July 13, 2015

Abstract
This appendix contains definitions of the distributions used throughout the paper, derivations of the full conditional distributions, and other details not included in the paper.

Keywords: state space models; filtering; Markov-switching; stochastic intensity; stochastic volatility; Bayesian inference, autoregressive-gamma process.
1 Definition of distributions

In this section, I define notation for the distributions that are used in the paper. The notation for the parameters of the distributions is local to this section of the appendix.

1.1 Normal gamma distribution

A normal gamma random variable $Y \sim N.G. (\mu, \alpha, \gamma, \nu)$ is obtained by taking a normal r.v. $Y \sim N(\mu + \gamma \sigma^2, \sigma^2)$ and allowing its variance to be a gamma r.v. $\sigma^2 \sim \text{Gamma}(\nu, 2\alpha^2 - \gamma^2)$. The p.d.f. of a N.G. random variable is

$$p(y|\mu, \alpha, \gamma, \nu) = \left(\frac{\alpha^2 - \gamma^2}{2}\right)^{\nu/2} \frac{\sqrt{2} \Gamma(\nu \alpha - \gamma)}{\sqrt{\pi} \Gamma(\nu)} \exp\left(\frac{\gamma^2}{\alpha^2 - \gamma^2}\right)$$

where $K_\lambda(x)$ is the modified Bessel function of the second kind; see, e.g. Abramowitz and Stegun (1964). The mean and variance of the distribution are

$$E[Y] = E[\sigma^2] E[Y|\sigma^2] = E[\sigma^2] (\mu + \gamma \sigma^2) = \mu + \frac{2\gamma \nu}{\alpha^2 - \gamma^2}$$

$$V[Y] = E[V(Y|\sigma^2)] + V[E(Y|\sigma^2)] = E[\sigma^2] + V[\mu + \gamma \sigma^2] = \frac{2\nu}{\alpha^2 - \gamma^2} + \frac{4\nu \gamma^2}{(\alpha^2 - \gamma^2)^2}$$

see, e.g. Kotz, Kozubowski, and Podgórski (2001). The parameter $\gamma$ controls the symmetry of the distribution such that $\gamma < 0$ is left-skewed, $\gamma > 0$ is right skewed, and a value of $\gamma = 0$ is symmetric.

1.2 Generalized inverse Gaussian distribution

A generalised inverse Gaussian random variable $X \sim \text{GIG}(\lambda, \chi, \psi)$ has p.d.f.

$$p(x|\lambda, \chi, \psi) = \left(\frac{\psi}{\chi}\right)^\lambda \frac{1}{2K_\lambda(\sqrt{\chi \psi})} x^{\lambda - 1} \exp\left(-\frac{1}{2} \left[\frac{\chi x^{-1}}{\psi} + \psi x\right]\right)$$

Moments of the GIG distribution are $E[X^\delta] = \left(\frac{\psi}{\chi}\right)^\delta \frac{K_{\lambda+\delta}(\sqrt{\chi \psi})}{K_\lambda(\sqrt{\chi \psi})}$. The c.d.f. of a GIG random variable needed to calculate quantiles can be expressed in terms of the incomplete Bessel function. Slevinsky and Safouhi (2010) provide a routine for its calculation that does not use numerical integration.

1.3 Sichel and negative binomial distributions

A random variable $Z \sim \text{Sichel}(\lambda, \chi, \psi)$ is obtained by taking a Poisson random variable $Z \sim \text{Poisson}(X)$ and allowing the mean $X$ to be a random draw from a GIG distribution $X \sim \text{GIG}(\lambda, \chi, \psi)$. The mass function for a Sichel random variable is

$$p(z|\lambda, \chi, \psi) = \left(\frac{\psi}{\psi + 2}\right)^\lambda \left(\frac{\chi}{\psi + 2}\right)^z \frac{1}{z!} \frac{K_{\lambda+z}(\sqrt{\chi (\psi + 2)})}{K_\lambda(\sqrt{\chi \psi})}, \quad z \geq 0$$

2
The first two moments of a Sichel distribution are often reported incorrectly in the literature. They follow from the law of iterated expectations as

\[ E[Z] = E_X [E(Z|X = x)] = E_X [X] = \left( \frac{\chi}{\psi} \right)^{\frac{1}{2}} \frac{K_{\lambda+1} \left( \sqrt{\chi \psi} \right)}{K_{\lambda} \left( \sqrt{\chi \psi} \right)} \]

\[ E[Z^2] = E_X [E(Z^2|X = x)] = E_X [X + X^2] = \left( \frac{\chi}{\psi} \right)^{\frac{3}{2}} \frac{K_{\lambda+1} \left( \sqrt{\chi \psi} \right)}{K_{\lambda} \left( \sqrt{\chi \psi} \right)} + \left( \frac{\chi}{\psi} \right)^{\frac{1}{2}} \frac{K_{\lambda+2} \left( \sqrt{\chi \psi} \right)}{K_{\lambda} \left( \sqrt{\chi \psi} \right)} \]

A negative binomial random variable \( K \sim \text{Neg. Bin.} (\omega, p) \) has p.m.f.

\[ p(k|\omega, p) = \frac{\Gamma(\omega + k)}{\Gamma(\omega) \Gamma(k+1)} p^k (1-p)^{\omega} \]

with mean \( E[K] = \frac{\omega}{1-p} \). The negative binomial distribution is a special case of the Sichel \((\lambda, \chi, \psi)\) as \( \chi \to 0 \).

1.4 Bessel distribution

A Bessel random variable \( S \sim \text{Bessel} (\kappa, \gamma) \) has probability mass function

\[ p(s|\kappa, \gamma) = \frac{1}{I_\kappa(\gamma) s! \Gamma(s+\kappa+1)} \left( \frac{\gamma}{2} \right)^{s+\kappa}, \quad s \geq 0 \]

where \( I_\kappa(\gamma) \) is the modified Bessel function of the first kind; see, e.g. Abramowitz and Stegun (1964). The mean and variance are

\[ E[S] = \frac{\gamma}{2} \frac{I_{\kappa+1}(\gamma)}{I_\kappa(\gamma)} \]

\[ V[S] = \gamma^2 \frac{I_{\kappa+2}(\gamma)}{I_\kappa(\gamma)} - \left( \frac{I_{\kappa+1}(\gamma)}{I_\kappa(\gamma)} \right)^2 \left( \frac{I_{\kappa+2}(\gamma)}{I_\kappa(\gamma)} - \frac{I_{\kappa+1}(\gamma)}{I_\kappa(\gamma)} \right) \]

1.5 Hypergeometric function distributions

Lemma 1.3.3 of Muirhead (1982) states that

\[ z^{-\alpha} \Gamma(a) \left( a; b; x \right) = \int_0^\infty t^{a-1} \exp(-zt) pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; kt) dt \]

where \( pF_q(a, b; x) \) is the generalized hypergeometric function; see, Mathai and Saxena (1973). This defines a family of probability distributions, where the left hand side is the normalizing constant.

1.6 Tricomi distribution

Another distribution for positive random variables \( h \) has the density

\[ p(h|\alpha, \beta, \psi) = \frac{1}{U(\alpha, \beta, \psi) \Gamma(\alpha)} h^{\alpha-1} (1+h)^{\beta-\alpha-1} \exp(-\psi h) \]

with \( \alpha > 0 \) and where \( U(\alpha, \beta, \psi) \) is the Tricomi confluent hypergeometric function. In this appendix, I will call this a Tricomi random variable. The distribution nests the gamma distribution when \( \beta = \alpha + 1 \). The function \( U(\alpha, \beta, \psi) \) is related to the standard confluent hypergeometric function \( _1F_1(a, b; x) \) through an identity.
1.7 Poisson-Tricomi distribution

A random variable $Z \sim \text{Poisson-Tricomi} (\alpha, \beta, \psi)$ is obtained by taking a Poisson random variable $Z \sim \text{Poisson}(\frac{H}{c})$ and allowing the mean $H$ to be a random draw from a continuous Tricomi distribution $H \sim \text{Tricomi}(\alpha, \beta, \psi)$. The mass function is

$$p(z|\alpha, \beta, \psi, c) = \int_0^\infty p(z|h) p(h) dh,$$

$$= \int_0^\infty \frac{1}{\Gamma(z+1)} \left(\frac{h}{c}\right)^z \exp \left(-\frac{h}{c}\right) U(\alpha, \beta, \psi) \Gamma(\alpha) h^{\alpha-1} (1+h)^{\beta-\alpha-1} \exp(-\psi h) dh,$$

$$= \frac{1}{U(\alpha, \beta, \psi) \Gamma(\alpha) \Gamma(z+1)} \int_0^\infty \left(\frac{1}{c}\right)^z h^{\alpha+z-1} (1+h)^{\beta-\alpha-1} \exp \left(-\frac{(c \psi + 1)}{c} h\right) dh,$$

$$= \frac{U(\alpha + z, \beta + z, \frac{(c \psi + 1)}{c}) \Gamma(\alpha + z)}{U(\alpha, \beta, \psi) \Gamma(\alpha) \Gamma(z+1)} \left(\frac{1}{c}\right)^z,$$

$z \geq 0$

2 Implementation

When implementing these procedures and evaluating the densities $p(y|z, x; \theta)$ and $p(z|z_{t-1}, y_{t-1}, x_{t-1}; \theta)$, all calculations should be performed in logarithms.

2.1 Evaluation of Bessel functions

For the stochastic volatility and duration models, the densities $p(y|z, x; \theta)$ and $p(z|z_{t-1}, y_{t-1}, x_{t-1}; \theta)$ are functions of the modified Bessel function of the second kind $K_{\nu+z}(x)$ where the order parameter is a function of the discrete state variable. In many software packages, this function cannot be evaluated “directly” for arbitrary values of $\nu + z$ and a fixed argument $x$.

Define the ratio of Bessel functions as $R_{\nu+j}(x) = \frac{K_{\nu+j}(x)}{K_{\nu+j-1}(x)}$. To calculate modified Bessel functions in a computationally stable manner, start by evaluating the function twice at stable values, e.g. $K_{\nu}(x)$ and $K_{\nu+1}(x)$, and calculate $R_{\nu+1}(x) = \exp[\log(K_{\nu+1}(x)) - \log(K_{\nu}(x))]$. Then, apply the recursion for $j = 2, \ldots,$

$$R_{\nu+j}(x) = \frac{1}{R_{\nu+j-1}(x)} + \frac{2}{x} (\nu + j - 1),$$

$$\log(K_{\nu+j}(x)) = \log(R_{\nu+j}(x)) + \log(K_{\nu+j-1}(x)),$$

which gives $K_{\nu+z}(x)$ for any $z$. These can be deduced from the well-known recursive formula $K_{\nu+j+1}(x) = K_{\nu+j-1}(x) + \frac{2(\nu+j)}{x} K_{\nu+j}(x)$.

2.2 Evaluation of Gamma functions

The Gamma function satisfies the recursive equation $\Gamma(\nu+1) = \nu \Gamma(\nu)$. For the stochastic count model, the expression $\Gamma(\nu + z_t)$ can be calculated in logarithms for any $z_t$ using this formula.
3 Models

In this appendix, I calculate the conditional likelihood \( p(y_t|z_t; \theta) \) and Markov transition distribution \( p(z_t|z_{t-1}, y_{t-1}; \theta) \) for the Cox Process (stochastic count), stochastic volatility, and stochastic duration examples.

3.1 Cox process (stochastic count) model

Let \( y_t = (y_{1t}, \ldots, y_{N_t}) \), \( \bar{y}_t = \sum_{i=1}^{N_t} y_{it} \), \( \hat{\beta}_t = \tau \exp \left( \sum_{i=1}^{N_t} y_{it} x_{it} \beta \right) \), and \( \beta_t = \tau \sum_{i=1}^{N_t} \exp (x_{it} \beta) \). The conditional likelihood obtained by integrating \( h_t \) out of the original measurement equation is

\[
p(y_t|z_t, x_t; \theta) = \int_0^\infty p(y_t|h_t, x_t; \theta)p(h_t|z_t; \theta)dh_t
\]

\[
= \int_0^\infty h_t^{\sum_{i=1}^{N_t} y_{it}} \exp \left( \sum_{i=1}^{N_t} y_{it} x_{it} \beta \right) \exp \left( -h_t \sum_{i=1}^{N_t} \exp (x_{it} \beta) \right) \frac{1}{\Gamma(\nu + z_t)} c^{-(\nu + z_t)} h_t^{(\nu + z_t) - 1} \exp \left( - \frac{h_t}{c} \right) dh_t
\]

\[
= \frac{\hat{\beta}_t}{\Gamma(\nu + z_t)} \left( \frac{1}{c} \right)^{\nu + z_t} \int_0^\infty h_t^{\nu + z_t + \bar{y}_t} \exp \left( - \frac{h_t (1 + c \hat{\beta}_t)}{c} \right) dh_t
\]

\[
= \frac{\Gamma(\nu + z_t + \bar{y}_t) \hat{\beta}_t}{\Gamma(\nu + z_t)} \left( \frac{1}{1 + c \hat{\beta}_t} \right)^{\nu + z_t} \left( \frac{1 + c \hat{\beta}_t}{1} \right) ^ {\nu + z_t}
\]

The Markov transition distribution for \( z_t \) is

\[
p(z_t|z_{t-1}, y_{t-1}, x_{t-1}; \theta) = \int_0^\infty p(z_t|h_{t-1}; \theta)p(h_{t-1}|y_{t-1}, z_{t-1}, x_{t-1}; \theta)dh_{t-1}
\]

\[
= \int_0^\infty \frac{1}{\Gamma(z_t + 1)} \left( \frac{\phi h_{t-1}}{c} \right)^{z_t} \exp \left( - \frac{\phi h_{t-1}}{c} \right) \frac{1}{\Gamma(\nu + \bar{y}_{t-1} + z_{t-1})} h_{t-1}^{\nu + \bar{y}_{t-1} + z_{t-1} - 1} \exp \left( - h_{t-1} \left( 1 + c \beta_{t-1} \right) \right) dh_{t-1}
\]

\[
= \frac{1}{\Gamma(\nu + \bar{y}_{t-1} + z_{t-1}) \Gamma(z_t + 1)} \left( \frac{1 + c \beta_{t-1}}{c} \right)^{\nu + \bar{y}_{t-1} + z_{t-1}} \left( \frac{\phi}{c} \right)^{z_t} \int_0^\infty h_{t-1}^{\nu + \bar{y}_{t-1} + z_{t-1} - 1} \exp \left( - h_{t-1} \left( 1 + \phi + c \beta_{t-1} \right) \right) dh_{t-1}
\]

\[
= \frac{\Gamma(\nu + \bar{y}_{t-1} + z_{t-1}) \Gamma(z_t + 1)}{\Gamma(\nu + \bar{y}_{t-1} + z_{t-1}) \Gamma(z_t + 1)} \left( \frac{1 + c \beta_{t-1}}{c} \right)^{\nu + \bar{y}_{t-1} + z_{t-1}} \left( \frac{\phi}{c} \right)^{z_t}
\]

\[
= \text{Neg. Bin.} \left( \nu + \bar{y}_{t-1} + z_{t-1}, \frac{\phi}{1 + \phi + c \beta_{t-1}} \right)
\]
3.2 Stochastic volatility model:

The conditional likelihood obtained by integrating the variance $h_t$ out of the original measurement equation is

$$p(y_t | z_t, x_t; \theta) = \int_0^\infty p(y_t | h_t, x_t; \theta) p(h_t | z_t; \theta) dh_t$$

$$= \int_0^\infty \frac{1}{\sqrt{2\pi}h_t^{-1/2}} \exp \left( -\frac{1}{2} (y_t - \mu - x_t \beta - \gamma h_t)^2 h_t^{-1} \right) \frac{c^{-(\nu+z_t)}}{\Gamma(\nu+z_t)} h_t^{-(\nu+z_t)-1} \exp \left( -\frac{h_t}{c} \right) dh_t$$

$$= \left( \frac{1}{c} \right)^{\nu+z_t} \sqrt{2} |y_t - \mu - x_t \beta|^{\nu+z_t-\frac{1}{2}} K_{\nu+z_t-\frac{1}{2}} \left( \sqrt{\frac{\nu}{\nu+\gamma^2}} |y_t - \mu - x_t \beta| \right) \exp \left( |y_t - \mu - x_t \beta|/\beta \right) \sqrt{\pi^\nu (\nu+z_t)} \left( \frac{\nu+\gamma^2}{\nu} \right)^{\nu+z_t-\frac{1}{2}}$$

$$= \text{Normal Gamma} \left( \mu + x_t \beta, \frac{2}{\nu+\gamma^2}, \gamma, \nu + z_t \right)$$

The transition distribution of the discrete mixing variable conditional on observing $y_{t-1}$ is given by

$$p(z_t | z_{t-1}, y_{t-1}, x_{t-1}; \theta) = \int_0^\infty p(z_t | h_{t-1}; \theta) p(h_{t-1} | z_{t-1}, y_{t-1}; \theta) dh_{t-1}$$

$$= \int_0^\infty \left( \frac{\phi h_{t-1}}{c} \right)^z \frac{1}{z^!} \exp \left( -\frac{\phi h_{t-1}}{c} \right) \left[ (y_{t-1} - \mu - x_{t-1} \beta)^2 \right]^{-(\nu+z_{t-1}-\frac{1}{2})} \left( \sqrt{\frac{y_{t-1} - \mu - x_{t-1} \beta}{c+\gamma^2}} \right)^{\nu+z_{t-1}-\frac{1}{2}}$$

$$= 2K_{\nu+z_{t-1}-\frac{1}{2}} \left( \sqrt{\frac{y_{t-1} - \mu - x_{t-1} \beta}{c+\gamma^2}} \right) h_{t-1}^{\nu+z_{t-1}-\frac{1}{2}} \exp \left( -\frac{1}{2} \left[ (y_{t-1} - \mu - x_{t-1} \beta)^2 h_{t-1}^{-1} + \frac{\gamma^2}{c} h_{t-1} \right] \right) dh_{t-1}$$

$$= \left( \frac{2}{\nu+\gamma^2} + \frac{\gamma^2}{c} \right)^{\nu+z_{t-1}-\frac{1}{2}} \left( \frac{\phi}{c} \right)^z \left[ (y_{t-1} - \mu - x_{t-1} \beta)^2 \right]^{\nu+z_{t-1}-\frac{1}{2}} \left( \frac{2(1+\phi)}{c} + \gamma^2 \right)^{\nu+z_{t-1}-\frac{1}{2}} \frac{1}{z^!} K_{\nu+z_{t-1}-\frac{1}{2}} \left( \sqrt{\frac{y_{t-1} - \mu - x_{t-1} \beta}{c+\gamma^2}} \right) \frac{\gamma^2}{c} \left( \frac{2}{\nu+\gamma^2} \right)^{\nu+z_{t-1}-\frac{1}{2}} \left( \frac{\phi}{c} \right)^z \left[ (y_{t-1} - \mu - x_{t-1} \beta)^2 \right]^{\nu+z_{t-1}-\frac{1}{2}} \left( \frac{2(1+\phi)}{c} + \gamma^2 \right)^{\nu+z_{t-1}-\frac{1}{2}}$$

$$= \text{Sichel} \left( \nu + z_{t-1} - \frac{1}{2}, \frac{\phi}{c} (y_{t-1} - \mu - x_{t-1} \beta)^2, \frac{2}{\nu+\gamma^2} \right)$$

The Markov transition distribution of $z_t$ is non-homogenous as it depends on the most recent observation $y_{t-1}$. 

6
3.3 Stochastic duration model:

The conditional likelihood is

\[
p(y_t | z_t, x_t; \theta) = \int_0^\infty p(y_t | h_t, x_t; \theta) p(h_t | z_t; \theta) dh_t
\]

\[
= \int_0^\infty \frac{(\alpha \beta)^{\alpha-1}}{\Gamma(\alpha \beta)} e^{-\alpha \beta} \frac{\alpha^\alpha}{\Gamma(\alpha)} \left( 1 + \frac{h_t}{\alpha} \right)^{\alpha+z_t} \frac{1}{\Gamma(\alpha + z_t)} \left( 1 + \frac{h_t}{\alpha} \right)^{\alpha+z_t} dh_t
\]

\[
= \frac{2^{\alpha-1} (\alpha \beta)^{\alpha-1} c^{-\alpha}}{\Gamma(\alpha)} \int_0^\infty h_t^{\alpha+z_t-1} \exp \left( - \frac{2y_t}{\alpha} h_t^{\alpha+z_t-1} + \frac{2}{c} h_t \right) dh_t
\]

The Markov transition distribution is

\[
p(z_t | z_{t-1}, y_{t-1}, x_{t-1}; \theta) = \int_0^\infty p(z_t | h_{t-1}; \theta) p(h_{t-1} | z_{t-1}, y_{t-1}, x_{t-1}; \theta) dh_{t-1}
\]

\[
= \int_0^\infty \frac{1}{z_t!} \left( \frac{\phi h_{t-1}}{c} \right)^{\alpha} \exp \left( - \frac{\phi h_{t-1}}{c} \right) \left( \sqrt{\frac{\alpha \beta}{cy_{t-1}}} \right)^{\alpha+z_{t-1}-1} \frac{1}{2} \left[ \frac{2y_{t-1}}{\alpha} h_{t-1} + \frac{2}{c} h_{t-1} \right] dh_{t-1}
\]

\[
= \frac{1}{z_t!} \left( \frac{\phi}{c} \right)^{\alpha} \left( \sqrt{\frac{\alpha \beta}{cy_{t-1}}} \right)^{\alpha+z_{t-1}-1} \frac{1}{2} \left[ \frac{2y_{t-1}}{\alpha} h_{t-1} + \frac{2(1 + \phi)}{c} h_{t-1} \right] dh_{t-1}
\]

\[
= \text{Sichel} \left( \nu + z_{t-1} - \alpha, \frac{\phi}{c} \left( \sqrt{\frac{\alpha \beta}{cy_{t-1}}} \right)^{\alpha+z_{t-1}-1} \right)
\]

3.4 Stochastic count model with over-dispersion

Consider the stochastic count model with observations having a negative-binomial distribution

\[
y_t \sim \text{Neg Bin} (\omega, p_t) \quad p_t = \frac{h_t}{1 + h_t}
\]

The observation density is

\[
p(y_t | h_t, \theta) = \frac{\Gamma(\omega + y_t)}{\Gamma(\omega) \Gamma(y_t + 1)} \left( \frac{h_t}{1 + h_t} \right)^{y_t} \left( 1 - \frac{h_t}{1 + h_t} \right)^{\omega}
\]
The conditional likelihood is

\[ p(y_t | z_{t-1}, y_{t-1}, \theta) = \int_0^\infty p(y_t | h_{t}, \theta) p(h_t | z_{t-1}, y_{t-1}, \theta) \, dh_t \]

\[ = \int_0^\infty \frac{\Gamma(\omega + y_t)}{\Gamma(y_t + 1)} \left( \frac{h_t}{1 + h_t} \right)^{y_t} \left( 1 - \frac{h_t}{1 + h_t} \right)^{\omega} \frac{1}{\Gamma(\nu + z_t)} h_t^{\nu+z_t-1} c^{-\nu+z_t} \exp \left( -\frac{h_t}{c} \right) \, dh_t \]

\[ = \frac{\Gamma(\omega + y_t)}{\Gamma(\omega) \Gamma(y_t + 1)} \frac{1}{\Gamma(\nu + z_t)} \left( \frac{1}{c} \right)^{\nu+z_t} U \left( y_t + \nu + z_t, \nu + z_t + 1 - \omega, \frac{1}{c} \right) \]

The Markov transition distribution is

\[ p(z_t | z_{t-1}, y_{t-1}, \theta) = \int_0^\infty p(z_t | h_{t-1}, \theta) p(h_{t-1} | z_{t-1}, y_{t-1}, \theta) \, dh_{t-1} \]

\[ = \int_0^\infty \frac{1}{z_t!} \left( \frac{\phi h_{t-1}}{c} \right)^{z_t} \exp \left( -\frac{\phi h_{t-1}}{c} \right) \frac{1}{(1 + h_{t-1})^{-(y_{t-1}+\omega-1)} h_{t-1}^{\nu+y_{t-1}+z_{t-1}-1}} \exp \left( -\frac{h_{t-1}}{c} \right) \, dh_{t-1} \]

\[ = \frac{1}{U(y_{t-1} + z_{t-1} + \nu, z_{t-1} + \nu + 1 - \omega, \frac{1}{c})} \frac{1}{z_t!} \left( \frac{\phi}{c} \right)^{z_t} \]

\[ = \int_0^\infty \frac{1}{(1 + h_{t-1})^{-(y_{t-1}+\omega-1)} h_{t-1}^{\nu+y_{t-1}+z_{t-1}-1}} \exp \left( -\frac{(1 + \phi) h_{t-1}}{c} \right) \, dh_{t-1} \]

\[ = \frac{U(y_{t-1} + z_{t-1} + \nu, z_{t-1} + \nu + 1 - \omega, \frac{1 + \phi}{c}) \Gamma(y_{t-1} + z_{t-1} + \nu)}{U(y_{t-1} + z_{t-1} + \nu, z_{t-1} + \nu + 1 - \omega, \frac{1}{c}) \Gamma(z_{t-1} + 1) \Gamma(y_{t-1} + 1) \Gamma(1 + \nu)} \left( \frac{\phi}{c} \right)^{z_t} \]

This is the mass function of a Poisson-Tricomi random variable, see section 1.7 above.

### 3.5 Stochastic binomial model

Observations \( y_t \) having a binomial distribution with time-varying probability \( p_t \) can be defined as

\[ y_t \sim \text{Binomial}(n, p_t) \quad p_t = \frac{h_t}{1 + h_t} \]

The Bernoulli distribution is the special case when \( n = 1 \). The observation density is

\[ p(y_t | h_t, \theta) = \frac{\Gamma(n + 1)}{\Gamma(y_t + 1) \Gamma(n - y_t + 1)} \left( \frac{h_t}{1 + h_t} \right)^{y_t} \left( 1 - \frac{h_t}{1 + h_t} \right)^{n - y_t} \]
The marginal filtering distribution conditional on process (stochastic count), stochastic volatility, and stochastic duration examples. In this appendix, I calculate the conditional distributions

\[ p(y_t|z_t, \theta) = \int_0^\infty p(y_t|h_t, \theta) p(h_t|z_t, \theta) \, dh_t \]

\[ = \int_0^\infty \frac{\Gamma(n + 1)}{\Gamma(y_t + 1)\Gamma(n - y_t + 1)} \left( \frac{h_t}{1 + h_t} \right)^{y_t} \left( 1 - \frac{h_t}{1 + h_t} \right)^{n-y_t} \frac{1}{\Gamma(\nu + z_t)} h_t^{\nu+z_t-1} c^{-\nu-z_t} \exp \left(-\frac{h_t}{c} \right) \, dh_t \]

\[ = \frac{\Gamma(n + 1)}{\Gamma(y_t + 1)\Gamma(n - y_t + 1)\Gamma(\nu + z_t)} \frac{1}{c} \int_0^\infty h_t^{y_t+\nu+z_t-1} (1 + h_t)^{-n} \exp \left(-\frac{h_t}{c} \right) \, dh_t \]

\[ = \frac{\Gamma(n + 1)}{\Gamma(y_t + 1)\Gamma(n - y_t + 1)\Gamma(\nu + z_t)} \frac{1}{c} U \left( y_t + \nu + z_t, t, \nu + z_t - n + 1, \frac{1}{c} \right) \]

The Markov transition distribution is

\[ p(z_{t-1}|z_{t-1}, y_{t-1}, \theta) = \int_0^\infty p(z_t|h_{t-1}, \theta) p(h_{t-1}|z_{t-1}, y_{t-1}, \theta) \, dh_{t-1} \]

\[ = \int_0^\infty \frac{1}{z_t!} \left( \frac{\phi h_{t-1}}{c} \right)^{z_t} \exp \left(-\frac{\phi h_{t-1}}{c} \right) \frac{(1 + h_{t-1})^{-(n-1)-1}}{\Gamma(y_t-1+\nu+y_t+z_t-1+1)} \frac{1}{c} \exp \left(-\frac{h_{t-1}}{c} \right) \, dh_{t-1} \]

\[ = \frac{1}{U \left( y_t-1+z_t-1+\nu, y_t-1+z_t-1+\nu+n-1, \frac{1}{c} \right) \Gamma(y_t-1+z_t-1+\nu)} \frac{1}{c} \frac{\phi^{z_t}}{z_t!} \]

\[ = \frac{1}{U \left( y_t-1+z_t-1+\nu, y_t-1+z_t-1+\nu+n-1, \frac{1}{c} \right) \Gamma(y_t-1+z_t-1+\nu)} \frac{\phi^{z_t}}{z_t!} \]

This is a Poisson-Student distribution.

4 Full conditional distributions

In this appendix, I calculate the conditional distributions \( p(h_t|y_{1:t}, z_{1:t}; \theta) \) and \( p(h_t|y_{1:T}, z_{1:T}; \theta) \) for the Cox process (stochastic count), stochastic volatility, and stochastic duration examples.

4.1 Cox process (stochastic count) model:

The marginal filtering distribution conditional on \( z_t \) is

\[ p(h_t|y_{1:t}, z_{1:t}, x_{1:t}; \theta) \propto p(y_t|h_t, x_t; \theta)p(h_t|z_t; \theta) \]

\[ \propto \frac{h_t^{\nu} \beta_t}{\Gamma(y_t + 1)} \frac{h_t^{\nu+z_t-1} c^{-\nu-z_t}}{\Gamma(\nu + z_t)} \exp \left(-\frac{h_t}{c} \right) \]

\[ = \text{Gamma} \left( \nu + \bar{y}_t + z_t, \frac{c}{1 + c\beta_t} \right) \]
The marginal filtering distribution of the variance is:

$$p(h_t|y_{1:T}, z_{1:T}, x_{1:T}; \theta) \propto p(y_t|h_t; \theta)p(h_t|z_t; \theta)p(z_{t+1}|h_t; \theta)$$

$$\propto h_t^{\varphi_t} \exp \left( -h_t \beta_t h_t^{(\nu+z_{t+1})-1} \exp \left( -\frac{h_t}{c} \right) h_t \exp \left( -\frac{h_t}{c} \right) \right)$$

$$\propto h_t^{(\nu+\varphi_t+z_t+z_{t+1})-1} \exp \left( -h_t \left[ \frac{1 + \phi + c\beta_t}{c} \right] \right)$$

$$= \text{Gamma} \left( \nu + \varphi_t + z_t + z_{t+1}, \frac{c}{1 + \phi + c\beta_t} \right)$$

The marginal smoothing distribution at intermediate time points conditional on $z_t$ is:

$$p(h_t|y_{1:T}, z_{1:T}, x_{1:T}; \theta) \propto p(y_t|h_t; \theta)p(h_t|z_t; \theta)$$

$$\propto \left( \frac{\phi h_0}{c} \right)^{\frac{z_t}{z_1}} \frac{1}{z_1!} \exp \left( -\frac{\phi h_0}{c} \right) \frac{1}{\Gamma(\nu) h_0^{\nu-1}} \left( \frac{1 - \phi}{c} \right)^{\nu} \exp \left( -\frac{h_0(1 - \phi)}{c} \right)$$

$$\propto \exp \left( -\frac{\phi h_0}{c} h_0^{\nu+z_t-1} \exp \left( -\frac{h_0}{c} \right) \right)$$

$$= \text{Gamma} \left( \nu + z_t, \frac{1}{z_1} \right)$$

4.2 Stochastic volatility model:

The marginal filtering distribution of the variance is:

$$p(h_t|z_{1:T}, y_{1:T}; \theta) \propto p(y_t|h_t; \theta)p(h_t|z_t; \theta)$$

$$\propto h_t^{-1/2} \exp \left( -\frac{1}{2} \left[ y_t - \mu - x_t \beta - \gamma h_t \right]^2 h_t^{-1} \right) \frac{1}{\Gamma(\nu+z_t)} h_t^{(\nu+z_t)-1} \exp \left( -\frac{h_t}{c} \right)$$

$$\propto h_t^{(\nu-1/2+z_t-1)} \exp \left( -\frac{1}{2} \left[ (y_t - \mu - x_t \beta)^2 h_t^{-1} + \left( \frac{2}{c} + \gamma^2 \right) h_t \right] \right)$$

$$= \text{GIG} \left( \nu + z_t - \frac{1}{2}, (y_t - \mu - x_t \beta)^2, \frac{2}{c} + \gamma^2 \right)$$

The marginal smoothing distribution for the variance is:

$$p(h_t|z_{1:T}, y_{1:T}; \theta) \propto p(y_t|h_t; \theta)p(h_t|z_t; \theta)p(z_{t+1}|h_t; \theta)$$

$$\propto h_t^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \left[ y_t - \mu - x_t \beta - \gamma h_t \right]^2 h_t^{-1} \right) \frac{1}{\Gamma(\nu+z_t)} h_t^{(\nu+z_t)-1} \exp \left( -\frac{h_t}{c} \right)$$

$$\propto h_t^{(\nu+z_t+z_{t+1}-1)-1} \exp \left( -\frac{1}{2} \left[ (y_t - \mu - x_t \beta)^2 h_t^{-1} + \left( \frac{2(1 + \phi)}{c} + \gamma^2 \right) h_t \right] \right)$$

$$= \text{GIG} \left( \nu + z_t + z_{t+1} - \frac{1}{2}, (y_t - \mu - x_t \beta)^2, \frac{2(1 + \phi)}{c} + \gamma^2 \right)$$
The marginal filtering distribution of \( h_t \) is

\[
p(h_t | z_{1:T} \cup y_{1:T}; \theta) \propto p(z_t | h_t; \theta) p(h_t; \theta) \\
\propto h_t^{-\alpha} \exp \left( -\frac{y_t}{\exp(x_t \beta)} h_t^{-1} \right) h_t^{\nu + z_t - 1} \exp \left( -\frac{h_t}{c} \right)
\]

\[
= \text{GIG} \left( \nu + z_t - \alpha, \frac{2 y_t}{\exp(x_t \beta)}, \frac{2(1 + \phi)}{c} \right)
\]

The marginal smoothing distribution for \( h_t \) is

\[
p(h_t | z_{1:T} \cup y_{1:T}; \theta) \propto p(y_t | h_t; \theta) p(h_t | z_{1:T} \cup y_{1:T}; \theta) \\
\propto h_t^{-\alpha} \exp \left( -\frac{y_t}{\exp(x_t \beta)} h_t^{-1} \right) h_t^{\nu + z_t - 1} \exp \left( -\frac{h_t}{c} \right)
\]

\[
= \text{GIG} \left( \nu + z_t + z_{t+1} - \alpha, \frac{2 y_t}{\exp(x_t \beta)}, \frac{2(1 + \phi)}{c} \right)
\]

The distribution for the initial condition is

\[
p(h_0 | y_{1:T} \cup z_{1:T} \cup x_{1:T}; \theta) \propto p(z_1 | h_0; \theta) p(h_0; \theta) \\
\propto \left( \frac{\phi h_0}{c} \right)^{z_1} \frac{1}{z_1!} \exp \left( -\frac{\phi h_0}{c} \right) \frac{1}{\Gamma(\nu)} h_0^{\nu - 1} \exp \left( -\frac{h_0}{c} \right) \left( \frac{c}{1 - \phi} \right)^{\nu} \exp \left( -\frac{h_0(1 - \phi)}{c} \right)
\]

\[
= \text{Gamma} (\nu + z_1, c)
\]

### 4.3 Stochastic duration model:

The full conditional filtering density is

\[
p(h_t | z_{1:T} \cup y_{1:T}; \theta) \propto p(y_t | h_t; \theta) p(h_t | z_{1:T} \cup y_{1:T}; \theta) \\
\propto h_t^{-\alpha} \exp \left( -\frac{y_t}{\exp(x_t \beta)} h_t^{-1} \right) h_t^{\nu + z_t - 1} \exp \left( -\frac{h_t}{c} \right)
\]

\[
= \text{GIG} \left( \nu + z_t - \alpha, \frac{2 y_t}{\exp(x_t \beta)}, \frac{2(1 + \phi)}{c} \right)
\]

### 4.4 Stochastic count model with over-dispersion

The full conditional filtering density is

\[
p(h_t | z_{1:T} \cup y_{1:T} \cup \theta) \propto p(y_t | h_t, \theta) p(h_t | z_{1:T} \cup \theta) \\
\propto \frac{\Gamma(\omega + y_t)}{\Gamma(\omega) \Gamma(y_t + 1)} \left( \frac{1}{1 + h_t} \right)^{y_t} \left( 1 - \frac{h_t}{1 + h_t} \right)^{\omega} \frac{1}{\Gamma(\nu + z_t)} h_t^{\nu + z_t - 1} \exp \left( -\frac{h_t}{c} \right)
\]

\[
= (1 + h_t)^{-(y_t + \omega - 1)} h_t^{\nu + y_t + z_t - 1} \exp \left( -\frac{h_t}{c} \right)
\]
This is the density of a positive Tricomi random variable as in section 1.6 above
\[
p(h_t|z_t, y_t, \theta) = \frac{(1 + h_t)^{-(y_t + \omega - 1) - 1} h_t^{\nu + y_t + z_t - 1}}{U(y_t + z_t + \nu, z_t + \omega + 1 - \omega, \frac{1}{\sigma}) \Gamma(y_t + z_t + \nu)} \exp \left(-\frac{h_t}{c}\right)
\]

The full conditional smoothing density is
\[
p(h_t|z_1:T, y_1:T, \theta) \propto p(y_t|h_t, \theta) p(h_t|z_t, \theta) p(z_{t+1}|h_t, \theta)
\]
\[
\times \frac{\Gamma(n + 1)}{\Gamma(y_t + 1) \Gamma(n - y_t + 1)} \left(\frac{h_t}{1 + h_t}\right)^{y_t} \left(1 - \frac{h_t}{1 + h_t}\right)^{n - y_t} \frac{1}{\Gamma(\nu + z_t)} h_t^{\nu + z_t - 1} c^{-\nu - z_t} \exp \left(-\frac{h_t}{c}\right)
\]
\[
\propto (1 + h_t)^{-(n-1)-1} h_t^{\nu+y_t+z_t-1} \exp \left(-\frac{(1 + \phi) h_t}{c}\right)
\]

This is also the kernel of a Tricomi random variable. The distribution for the initial condition is a gamma distribution just like the other models \(h_0 \sim \text{Gamma}(\nu + z_1, c)\).

### 4.5 Stochastic binomial model

The full conditional filtering density is
\[
p(h_t|z_{1:t}, y_1:t, \theta) \propto p(y_t|h_t, \theta) p(h_t|z_t, \theta)
\]
\[
\times \frac{\Gamma(n + 1)}{\Gamma(y_t + 1) \Gamma(n - y_t + 1)} \left(\frac{h_t}{1 + h_t}\right)^{y_t} \left(1 - \frac{h_t}{1 + h_t}\right)^{n - y_t} \frac{1}{\Gamma(\nu + z_t)} h_t^{\nu + z_t - 1} c^{-\nu - z_t} \exp \left(-\frac{h_t}{c}\right)
\]
\[
\propto (1 + h_t)^{-(n-1)-1} h_t^{\nu+y_t+z_t-1} \exp \left(-\frac{h_t}{c}\right)
\]

This is the kernel of a Tricomi random variable
\[
p(h_t|z_t, y_t, \theta) = \frac{(1 + h_t)^{-(n-1)-1} h_t^{\nu+y_t+z_t-1}}{U(y_t + z_t + \nu, z_t + \nu + n - 1, \frac{1}{\sigma}) \Gamma(y_t + z_t + \nu)} \exp \left(-\frac{h_t}{c}\right)
\]

The full conditional smoothing density is
\[
p(h_t|z_1:T, y_1:T, \theta) \propto p(y_t|h_t, \theta) p(h_t|z_t, \theta) p(z_{t+1}|h_t, \theta)
\]
\[
\times \frac{\Gamma(n + 1)}{\Gamma(y_t + 1) \Gamma(n - y_t + 1)} \left(\frac{h_t}{1 + h_t}\right)^{y_t} \left(1 - \frac{h_t}{1 + h_t}\right)^{n - y_t} \frac{1}{\Gamma(\nu + z_t)} h_t^{\nu + z_t - 1} c^{-\nu - z_t} \exp \left(-\frac{h_t}{c}\right)
\]
\[
\propto (1 + h_t)^{-(n-1)-1} h_t^{\nu+y_t+z_t+1+z_t-1} \exp \left(-\frac{(1 + \phi) h_t}{c}\right)
\]

which also has a Tricomi kernel. The distribution for the initial condition is a gamma distribution just like the other models \(h_0 \sim \text{Gamma}(\nu + z_1, c)\).
5 Initial conditions

In this appendix, I calculate the initial likelihood \( p(y_1; \theta) \), filtering \( p(z_1|y_1; \theta) \), and predictive distributions for the stochastic count, stochastic volatility, and stochastic duration examples. The stationary distribution is a gamma distribution \( h_0 \sim \text{Ga}(\nu, \frac{c}{\nu}) \). One iteration of the Markov transition kernel leaves the distribution invariant such that \( h_1 \sim \text{Ga}(\nu, \frac{c}{\nu}) \).

5.1 Cox process (stochastic count) model:

The first contribution to the log-likelihood is

\[
p(y_1; \theta) = \int_0^\infty p(y_1|h_1; \theta)p(h_1; \theta)dh_1
\]

\[
= \int_0^\infty h_1^{\nu_0} \beta_1 \exp(-h_1 \beta_1) \left(1 - \phi \right) c \nu \frac{1}{\Gamma(\nu)} h_1^{\nu-1} \exp \left(-h_1 \left(1 - \phi \right) \frac{c}{\nu} \right) dh_1
\]

\[
= \frac{\beta_1}{\Gamma(\nu)} \left( \frac{1 - \phi}{c} \right) \int_0^\infty h_1^{\nu_0} \exp \left(-h_1 \left(1 - \phi \right) + c \beta_1 \right) dh_1
\]

\[
= \frac{\beta_1 \Gamma(\nu + \tilde{y}_1)}{\Gamma(\nu)} \left( \frac{1 - \phi}{c} \right)^\nu \left( \frac{c}{1 - \phi + c \beta_1} \right)^{\nu_0 + \tilde{y}_1}
\]

The filtering distribution \( p(h_1|y_1, x_1; \theta) \) is

\[
p(h_1|y_1, x_1; \theta) \propto p(y_1|h_1, x_1; \theta)p(h_1; \theta)
\]

\[
= h_1^{\nu_0 + \tilde{y}_1 - 1} \exp \left(-h_1 \left(1 - \phi \right) + c \beta_1 \right)
\]

\[
= \text{Gamma} \left( \nu + \tilde{y}_1, \frac{c}{1 - \phi + c \beta_1} \right)
\]

The filtering distribution \( p(z_1|y_1, x_1; \theta) \) is

\[
p(z_1|y_1, x_1; \theta) \propto p(y_1|z_1, x_1; \theta)p(z_1; \theta)
\]

\[
= \frac{\Gamma(\nu + z_1 + \tilde{y}_1)}{\Gamma(\nu + z_1)} \left( \frac{1}{1 + c \beta_1} \right)^{\nu + z_1} \left( \frac{c}{1 + c \beta_1} \right)^{\tilde{y}_1} \frac{1}{z_1!} \phi^{z_1}
\]

The one step ahead predictive distribution \( p(z_2|y_1, x_1; \theta) \) is

\[
p(z_2|y_1, x_1; \theta) = \int_0^\infty p(z_2|h_1; \theta)p(h_1|y_1, x_1; \theta)dh_1
\]

\[
= \frac{1}{\Gamma(\nu + \tilde{y}_1)} \left( \frac{1 - \phi}{c} + c \beta_1 \right)^\nu \frac{\phi^{z_2}}{z_2!} \int_0^\infty h_1^{\nu_0 + \tilde{y}_1 + z_2 - 1} \exp \left(-h_1 \left(1 + c \beta_1 \right) \frac{c}{\nu} \right) dh_1
\]

\[
= \frac{\Gamma(\nu + \tilde{y}_1 + z_2)}{\Gamma(\nu + \tilde{y}_1) z_2!} \left( \frac{1 - \phi}{c} + c \beta_1 \right)^\nu \frac{\phi^{z_2}}{z_2!} \left( \frac{c}{1 + c \beta_1} \right)^{\nu + z_2}
\]

\[
= \text{Neg. Bin} \left( \nu + \tilde{y}_1, \frac{\phi}{1 + c \beta_1} \right)
\]
The one step ahead predictive distribution $p(h_2|y_1, x_1; \theta)$ is

$$p(h_2|y_1, x_1; \theta) = \sum_{z_2=0}^{\infty} p(h_2|z_2; \theta)p(z_2|y_1, x_1; \theta)$$

$$= \sum_{z_2=0}^{\infty} \frac{1}{\Gamma(\nu + z_2)} h_2^{\nu + z_2 - 1} \left(\frac{1}{c}\right)^{\nu + z_2} \exp\left(-\frac{h_2}{c}\right)$$

$$\frac{\Gamma(\nu + y_1 + z_2)}{\Gamma(\nu + y_1) z_2!} \left(\frac{1}{\nu} - \phi + c \beta_1\right)^{\nu} \left(\frac{\theta}{c}\right)^{\nu} \left(\frac{1 + c \beta_1}{1 + c \beta_1}\right)^{\nu + z_2}$$

$$= \frac{1}{\Gamma(\nu + y_1) 2} \left(\frac{1}{c}\right)^{\nu} \exp\left(-\frac{h_2}{c}\right) \left(\frac{1}{\nu} + c \beta_1\right)^{\nu} \sum_{z_2=0}^{\infty} \frac{\Gamma(\nu + y_1 + z_2)}{\Gamma(\nu + z_2) z_2!} \left(\frac{\phi h_2}{c(1 + c \beta_1)}\right)^{\nu + z_2}$$

where $1 F_1 (a, b; x)$ is the (confluent) hypergeometric function. The likelihood for the second time period is

$$p(y_2|y_1, x_1; \theta) = \int_0^{\infty} p(h_2|y_2, x_2; \theta)p(h_2|y_1, x_1; \theta)dh_2$$

$$= \frac{1}{\Gamma(\nu + y_1) 2} \left(\frac{1}{c}\right)^{\nu} \exp\left(-\frac{h_2}{c}\right) \left(\frac{1}{\nu} + c \beta_1\right)^{\nu} \sum_{z_2=0}^{\infty} \frac{\Gamma(\nu + y_1 + z_2)}{\Gamma(\nu + z_2) z_2!} \left(\frac{\phi h_2}{c(1 + c \beta_1)}\right)^{\nu + z_2}$$

$$= \frac{1}{\Gamma(\nu + y_1) 2} \left(\frac{1}{c}\right)^{\nu} \exp\left(-\frac{h_2}{c}\right) \left(\frac{1}{\nu} + c \beta_1\right)^{\nu} \frac{2 F_1 \left(\nu + y_2, \nu + y_1, \nu; \frac{\phi h_2}{c(1 + c \beta_1)}\right)}{\nu + y_2}\right)$$

where $2 F_1 (a_1, a_2, b; x)$ is the Gaussian hypergeometric function. The second filtering distribution is

$$p(h_2|y_1, x_1; \theta) = \frac{p(y_2|y_1, x_1; \theta)p(h_2|y_1, x_1; \theta)}{p(y_1|y_1, x_1; \theta)}$$

$$= \frac{1}{\Gamma(\nu + y_1)} 2 \left(\frac{1}{c}\right)^{\nu} \exp\left(-\frac{h_2}{c}\right) \left(\frac{1}{\nu} + c \beta_1\right)^{\nu} \frac{2 F_1 \left(\nu + y_2, \nu + y_1, \nu; \frac{\phi h_2}{c(1 + c \beta_1)}\right)}{\nu + y_2}\right)$$

$$= \frac{1}{\Gamma(\nu + y_1)} 2 \left(\frac{1}{c}\right)^{\nu} \exp\left(-\frac{h_2}{c}\right) \left(\frac{1}{\nu} + c \beta_1\right)^{\nu} \frac{2 F_1 \left(\nu + y_2, \nu + y_1, \nu; \frac{\phi h_2}{c(1 + c \beta_1)}\right)}{\nu + y_2}\right)$$

$$= \frac{1}{\Gamma(\nu + y_1)} 2 \left(\frac{1}{c}\right)^{\nu} \exp\left(-\frac{h_2}{c}\right) \left(\frac{1}{\nu} + c \beta_1\right)^{\nu} \frac{2 F_1 \left(\nu + y_2, \nu + y_1, \nu; \frac{\phi h_2}{c(1 + c \beta_1)}\right)}{\nu + y_2}\right)$$

$$= \frac{1}{\Gamma(\nu + y_1)} 2 \left(\frac{1}{c}\right)^{\nu} \exp\left(-\frac{h_2}{c}\right) \left(\frac{1}{\nu} + c \beta_1\right)^{\nu} \frac{2 F_1 \left(\nu + y_2, \nu + y_1, \nu; \frac{\phi h_2}{c(1 + c \beta_1)}\right)}{\nu + y_2}\right)$$
The third predictive distribution \( p(z_3 | y_{1:2}, x_{1:2}; \theta) \) is

\[
p(z_3 | y_{1:2}, x_{1:2}; \theta) = \int_0^\infty p(z_3 | h_2; \theta) p(h_2 | y_{1:2}, x_{1:2}; \theta) dh_2
\]

\[
= \frac{1}{\Gamma(\nu + \bar{y}_2)} \left( \frac{1 + c\beta_2}{c} \right)^{\nu + \bar{y}_2} \frac{1}{2} F_1 \left( \nu + \bar{y}_2, \nu + \bar{y}_1, \nu; \frac{\phi}{(1 + c\beta_2)(1 + c\beta_1)} \right) \frac{1}{z_3} \left( \frac{\phi}{c} \right)^{z_3}
\]

\[
\int_0^\infty h_2^{\nu + \bar{y}_2 + z_3 - 1} \exp \left( -h_2 \frac{(1 + \phi + c\beta_2)}{c} \right) 2 F_1 \left( \nu + \bar{y}_2, \nu + \bar{y}_1, \nu; \frac{\phi h_2}{c(1 + c\beta_1)} \right) dh_2
\]

\[
= \frac{1}{\Gamma(\nu + \bar{y}_2)} \frac{1}{\Gamma(\nu)} \left( \frac{1 + c\beta_2}{c} \right)^{\nu + \bar{y}_2} \frac{1}{2} F_1 \left( \nu + \bar{y}_2, \nu + \bar{y}_1, \nu; \frac{\phi}{(1 + c\beta_2)(1 + c\beta_1)} \right) \frac{1}{z_3} \left( \frac{\phi}{c} \right)^{z_3}
\]

\[
\left( \frac{c}{1 + \phi + c\beta_2} \right)^{\nu + \bar{y}_2 + z_3} 2 F_1 \left( \nu + \bar{y}_2 + z_3, \nu + \bar{y}_1, \nu; \frac{\phi}{(1 + c\beta_1)(1 + \phi + c\beta_2)} \right)
\]

where I have used Lemma 1.3.3 of Muirhead (1982).

### 5.2 Stochastic volatility model:

The first contribution to the log-likelihood is

\[
p(y_1 | x_1; \theta) = \int_0^\infty p(y_1 | h_1; \theta) p(h_1; \theta) dh_1
\]

\[
= \int_0^\infty \frac{1}{\sqrt{2\pi}} h_1^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (y_1 - \mu - x_1 \beta h_1)^2 h_1^{-1} \right) \left( \frac{c}{1 - \phi} \right)^{-\nu} \frac{1}{\Gamma(\nu)} h_1^{\nu - 1} \exp \left( -\frac{h_1(1 - \phi)}{c} \right) dh_1
\]

\[
= \left( \frac{1 - \phi}{c} \right)^{\nu} \frac{1}{\Gamma(\nu)} \sqrt{2\pi} \left[ (y_1 - \mu - x_1 \beta)^2 h_1^{-1} + \left( \frac{2(1 - \phi)}{c} + \gamma^2 \right) h_1^{-1} \right] \exp \left( \gamma (y_1 - \mu - x_1 \beta) \right) \int_0^\infty h_1^{\nu - 1} \exp \left( -\frac{h_1(1 - \phi)}{c} \right) dh_1
\]

\[
= \left( \frac{1 - \phi}{c} \right)^{\nu} \sqrt{2} (y_1 - \mu - x_1 \beta)^{\nu - \frac{1}{2}} K_{\nu - \frac{1}{2}} \left( \sqrt{\frac{2(1 - \phi)}{c} + \gamma^2} \right) (y_1 - \mu - x_1 \beta) \exp \left( \gamma (y_1 - \mu - x_1 \beta) \right)
\]

The filtering distribution \( p(h_1 | y_1, x_1; \theta) \) follows from Bayes’ rule

\[
p(h_1 | y_1; \theta) \propto p(y_1 | h_1; \theta) p(h_1; \theta)
\]

\[
\propto h_1^{-1/2} \exp \left( -\frac{1}{2} (y_1 - \mu - x_1 \beta)^2 h_1^{-1} + \gamma^2 h_1 \right) \exp \left( -\frac{h_1(1 - \phi)}{c} \right) h_1^{\nu - 1}
\]

\[
\propto h_1^{\nu - 1 - 1} \exp \left( -\frac{1}{2} (y_1 - \mu - x_1 \beta)^2 h_1^{-1} + \left( \frac{2(1 - \phi)}{c} + \gamma^2 \right) h_1 \right)
\]

\[
= \text{GIG} \left( \nu - \frac{1}{2}, (y_1 - \mu - x_1 \beta)^2, \frac{2(1 - \phi)}{c} + \gamma^2 \right)
\]
The conditional filtering distribution of \( z_1 \) is

\[
p(z_1|y_1, x_1; \theta) \propto p(y_1|z_1, x_1; \theta)p(z_1; \theta)
\]

\[
\propto \frac{1}{\Gamma(\nu + z_1)} \left( \frac{\phi}{c} \right)^{z_1} \frac{K_{\nu+z_1-\frac{1}{2}} \left( \sqrt{\frac{2}{c}} + \gamma^2 |y_1 - \mu - x_1\beta| \right)}{\left( |y_1 - \mu - x_1\beta| \right)^{\nu + 1} \Gamma(\nu + z_1)} \frac{1}{z_1!} \Gamma(\nu + z_1)
\]

\[
\propto \left( \frac{\phi}{c} \right)^{z_1} \frac{K_{\nu+z_1-\frac{1}{2}} \left( \sqrt{\frac{2}{c}} + \gamma^2 |y_1 - \mu - x_1\beta| \right)}{z_1!} \frac{1}{z_1!} \Gamma(\nu + z_1)
\]

\[
= \text{Sichel} \left( \nu - 1, \frac{1}{2}, \frac{\phi}{c} (|y_1 - \mu - x_1\beta|^2)^{\frac{1}{2}}, \frac{c}{\phi} \left[ \frac{2(1 - \phi)}{c} + \gamma^2 \right] \right)
\]

A Sichel distribution is typically derived as a Poisson random variable with a GIG mixing distribution. Here, it is the posterior distribution of a normal gamma likelihood combined with a negative binomial prior. The one-step ahead predictive distribution of \( z_2 \) is

\[
p(z_2|y_1, x_1; \theta) = \int_0^\infty p(z_2|h_1; \theta)p(h_1|y_1; \theta)dh_1
\]

\[
= \int_0^\infty \frac{1}{\tilde{z}_2!} \left( \frac{\phi h_1}{c} \right)^{\tilde{z}_2} \exp \left( -\frac{\phi h_1}{c} \right)
\]

\[
\left[ (y_1 - \mu)^2 \right]^{-(\nu - \frac{1}{2})} \left[ (y_1 - \mu)^{\frac{2(1 - \phi)}{c} + \gamma^2} \right]^{\nu - \frac{1}{2}}
\]

\[
\frac{2K_{\nu-\frac{1}{2}} \left( |y_1 - \mu|^{\frac{2(1 - \phi)}{c} + \gamma^2} \right)}{h_1^{\nu - \frac{1}{2} - 1}} \exp \left( -\frac{1}{2} \left( (y_1 - \mu)^2 h_1^{-1} + \left( \frac{2(1 - \phi)}{c} + \gamma^2 \right) h_1 \right) \right) dh_1
\]

\[
= \left( \frac{2(1 - \phi)}{c} + \gamma^2 \right)^{\frac{\nu - \frac{1}{2}}{2}} \left( \frac{\phi}{c} \sqrt{\frac{2}{c} + \gamma^2} \right)^{z_2} \frac{1}{z_2!} \frac{K_{\nu+z_2-\frac{1}{2}} \left( |y_1 - \mu - x_1\beta|^{\frac{2}{c} + \gamma^2} \right)}{z_2!} \frac{1}{z_2!} \Gamma(\nu + z_2)
\]

\[
= \text{Sichel} \left( \nu - 1, \frac{1}{2}, \frac{\phi}{c} (|y_1 - \mu - x_1\beta|^2)^{\frac{1}{2}}, \frac{c}{\phi} \left[ \frac{2(1 - \phi)}{c} + \gamma^2 \right] \right)
\]

This is the usual construction of a Sichel distribution as a GIG mixture of Poisson random variables.
5.3 Stochastic duration model:

The first contribution to the log-likelihood is

\[
p(y_1|x_1; \theta) = \int_0^\infty p(y_1|h_1, x_1; \theta)p(h_1; \theta)dh_1
\]

\[
= \int_0^\infty \frac{\exp(x_1 \beta)^{-\alpha}}{\Gamma(\alpha)} h_1^{-\alpha} \exp\left(-\frac{y_1}{\exp(x_1 \beta)} h_1^{-1}\right) \left(\frac{1 - \phi}{c}\right)^\nu \frac{1}{\Gamma(\nu)} h_1^{\nu - 1} \exp\left(-\frac{h_1(1 - \phi)}{c}\right) dh_1
\]

\[
= \frac{\exp(x_1 \beta)^{-\alpha}}{\Gamma(\alpha) \Gamma(\nu)} \left(\frac{1 - \phi}{c}\right)^\nu \int_0^\infty h_1^{\nu - \alpha - 1} \exp\left(-\frac{1}{2} \left[\frac{2y_1}{\exp(x_1 \beta)} h_1^{-1} + \frac{2(1 - \phi)}{c} h_1\right]\right) dh_1
\]

\[
= \frac{2 \exp(x_1 \beta)^{-\alpha}}{\Gamma(\alpha) \Gamma(\nu)} \left(\frac{1 - \phi}{c}\right)^\nu \left(\frac{\frac{2y_1}{\exp(x_1 \beta)}}{2(1 - \phi)}\right)^{\nu - \alpha} K_{\nu - \alpha}\left(\frac{\frac{4y_1(1 - \phi)}{c \exp(x_1 \beta)}}{2(1 - \phi)}\right).
\]

This density does not apparently have a name. The filtering distribution \(p(h_1|y_1, x_1; \theta)\) follows from Bayes’ rule

\[
p(h_1|y_1, x_1; \theta) \propto p(y_1|h_1; \theta)p(h_1; \theta)
\]

\[
= h_1^{-\alpha} \exp\left(-\frac{y_1}{\exp(x_1 \beta)} h_1^{-1}\right) h_1^{\nu - 1} \exp\left(-\frac{h_1(1 - \phi)}{c}\right).
\]

The conditional filtering distribution of \(z_1\) is

\[
p(z_1|x_1; \theta) \propto p(y_1|z_1; \theta)p(z_1; \theta)
\]

\[
= c^{-z_1} \frac{\sqrt{y_1 c}}{z!} \binom{z_1}{\nu - \alpha} K_{\nu + z_1 - \alpha}\left(\sqrt{\frac{4y_1}{c \exp(x_1 \beta)}}\right).
\]

The one-step ahead predictive distribution of \(z_2\) is

\[
p(z_2|y_1, x_1; \theta) = \int_0^\infty p(z_2|h_1; \theta)p(h_1|y_1, x_1; \theta)dh_1
\]

\[
= \text{Sichel}\left(\nu - \alpha, \frac{\phi}{c \exp(x_1 \beta)}, \frac{y_1}{2(1 - \phi)}\right).
\]

17
5.4 Stochastic count model with over-dispersion

The initial contribution to the likelihood is

\[
p(y_1|\theta) = \int_0^\infty p(y_1|h_1, \theta) p(h_1|\theta) dh_1
\]

\[
= \int_0^\infty \frac{\Gamma(\omega + y_1)}{\Gamma(\omega) \Gamma(y_1 + 1)} \left( \frac{h_1}{1 + h_1} \right)^y \left( \frac{1}{1 + h_1} \right)^{\omega} \left( \frac{1 - \frac{\phi}{c}}{c} \right)^{\nu} \frac{1}{\Gamma(\nu)} h_1^{\nu - 1} \exp \left( -\frac{h_1(1 - \phi)}{c} \right) dh_1
\]

\[
= \frac{\Gamma(\omega + y_1) \Gamma(y_1 + 1)}{\Gamma(\omega) \Gamma(y_1 + 1)} \left( \frac{1 - \frac{\phi}{c}}{c} \right)^{\nu} \frac{1}{\Gamma(\nu)} \int_0^\infty (1 + h_1)^{-(\omega + y_1 - 1)} h_1^{\nu + y_1 - 1} \exp \left( -\frac{h_1(1 - \phi)}{c} \right) dh_1
\]

\[
= \frac{\Gamma(\omega + y_1) \Gamma(y_1 + 1) U(\nu + y_1, \nu + 1 - \omega, \frac{1 - \phi}{c})}{\Gamma(\nu)} \left( \frac{1 - \frac{\phi}{c}}{c} \right)^{\nu} \frac{1}{\Gamma(\nu)}
\]

The initial filtering distribution has the density of a Tricomi random variable

\[
p(h_1|y_1, \theta) = \frac{p(y_1|h_1, \theta) p(h_1|\theta)}{p(y_1|\theta)}
\]

\[
= \frac{1}{p(y_1|\theta)} \frac{\Gamma(\omega + y_1)}{\Gamma(\omega) \Gamma(y_1 + 1)} \left( \frac{h_1}{1 + h_1} \right)^y \left( \frac{1}{1 + h_1} \right)^{\omega} \left( \frac{1 - \frac{\phi}{c}}{c} \right)^{\nu} \frac{1}{\Gamma(\nu)} h_1^{\nu - 1} \exp \left( -\frac{h_1(1 - \phi)}{c} \right)
\]

\[
= \frac{1}{\Gamma(\omega) \Gamma(y_1 + 1) U(\nu + y_1, \nu + 1 - \omega, \frac{1 - \phi}{c})} \left( \frac{1 + h_1}{1} \right)^{-(\omega + y_1 - 1)} h_1^{\nu + y_1 - 1} \exp \left( -\frac{h_1(1 - \phi)}{c} \right)
\]

The predictive distribution for the auxiliary variable for date two is

\[
p(z_2|y_1, \theta) = \int_0^\infty p(z_2|h_1, \theta) p(h_1|y_1, \theta) dh_1
\]

\[
= \int_0^\infty \frac{1}{\Gamma(z_2 + 1)} \left( \frac{\phi h_1}{c} \right)^{z_2} \exp \left( -\frac{\phi h_1}{c} \right) \frac{1}{\Gamma(\nu + y_1) U(\nu + y_1, \nu + 1 - \omega, \frac{1 - \phi}{c})} \left( \frac{1 + h_1}{1} \right)^{-(\omega + y_1 - 1)} h_1^{\nu + y_1 - 1} \exp \left( -\frac{h_1(1 - \phi)}{c} \right) dh_1
\]

\[
= \frac{\Gamma(\omega + y_1) \Gamma(\nu + y_1) U(\nu + y_1, \nu + 1 - \omega, \frac{1 - \phi}{c})}{\Gamma(z_2 + 1) \Gamma(\nu + y_1) U(\nu + y_1, \nu + 1 - \omega, \frac{1 - \phi}{c})} \left( \frac{\phi}{c} \right)^{z_2} \frac{1}{\Gamma(\nu + y_1)} U(\nu + y_1, \nu + 1 - \omega, \frac{1 - \phi}{c})
\]

\[
= \frac{\Gamma(\nu + z_2 + y_1) U(\nu + z_2 + y_1, \nu + z_2 + 1 - \omega, \frac{1 - \phi}{c})}{\Gamma(z_2 + 1) \Gamma(\nu + y_1) U(\nu + y_1, \nu + 1 - \omega, \frac{1 - \phi}{c})} \left( \frac{\phi}{c} \right)^{z_2}
\]

6 Particle filters

The following is a longer discussion of Section 4 of the paper, which compares particle filters.
6.1 Filtering recursions based on particle filters

An alternative method for calculating the likelihood function and filtering distributions for nonlinear, non-Gaussian state space models are sequential Monte Carlo methods also known as particle filters; see, e.g. Doucet and Johansen (2011) and Creal (2012). Particle filters approximate distributions whose support is infinite by a finite set of points and probability masses. The finite state Markov switching algorithm can be interpreted as a particle filter with \( Z + 1 \) particles. The primary difference between a particle filter and the approach taken here is a standard particle filter selects the support points for \( z_t \) at random via Monte Carlo draws while the algorithms of the paper use a deterministic set of points of support, i.e. the integers from 0 to \( Z \). In the terminology of the particle filtering literature, I have used Rao-Blackwellisation to analytically integrate out the continuous-valued state variable \( h_t \); see e.g. Chen and Liu (2000) and Chopin (2004). Then, in the prediction step, the Markov-switching algorithm marginalizes out the discrete state variable by transitioning \( z_{t-1} \) through all of the possible future \( Z \) states, e.g. Klass, de Freitas, and Doucet (2005).

In this section, I compare the log-likelihood functions calculated from the new algorithm with a standard particle filtering algorithm to demonstrate their relative accuracy. All filtering algorithms were run with the parameter values fixed at the ML estimates from the S&P 500 dataset. I run the Markov-switching algorithm with the truncation parameter set at \( Z = 3500 \).

For comparison purposes, I implement a particle filter with the transition density as a proposal. The particles are resampled at random times according to the effective sample size (ESS) using residual resampling, see Liu and Chen (1998) and Creal (2012). This is a simple extension of the original particle filter of Gordon, Salmond, and Smith (1993). I also implemented the auxiliary particle filter of Pitt and Shephard (1999) and the results were practically identical.

To compare the methods, I compute slices of the log-likelihood function for the parameters \( \phi \) and \( \nu \) over the regions \([0.97, 0.999]\) and \([1.0, 2.2]\), respectively. Cuts of the log-likelihood function for other parameters \((\mu, \gamma, c)\) are similar and are available in an online appendix. Each of these regions are divided into 1000 equally spaced points, where the log-likelihood function is evaluated. While the values of \( \phi \) and \( \nu \) change one at a time, the remaining parameters are held fixed at their ML estimates. The log-likelihood functions from the particle filter are shown for different numbers of particles \( N = 10000 \) and \( 30000 \), which are representative of values used in the literature. As seen from Figure 3, the particle filter does not produce an estimate of the log-likelihood function that is smooth in the parameter space. Consequently, standard derivative-based optimization routines have trouble converging to the maximum. The smoothness of the log-likelihood function produced by the new approach is an attractive feature that makes
Figure 1: Slices of the log-likelihood function for $\phi$ (left) and $\nu$ (right) for the new algorithm (blue solid line) versus the particle filter (red dots). The particle sizes are $N = 10000$ (top row) and $N = 30000$ (bottom row). The vertical (green) lines are the ML estimates of the model reported in Table ??.
calculation of the ML estimates straightforward. In this application, there is also a sizeable downward bias in the particle filter’s estimates of the log-likelihood function and a reasonably large amount of Monte Carlo variability. The particle filter’s estimator of the likelihood function is unbiased as well as consistent and asymptotically normal; see, e.g. Chapter 11 of Del Moral (2004). Taking logarithms causes the particle filter’s estimator to be biased downward due to Jensen’s inequality.

6.2 Additional cuts of the log-likelihood

This section includes slices of the log-likelihood function for the parameters $\mu, c$ and $\beta$ over the regions $[0, 0.2]$, $[0.01, 0.02]$, and $[-0.15, 0.05]$, respectively. The particle filter was run for $N = 10000, 30000$, and $100000$. A similar Monte Carlo experiment was performed with other particle filtering algorithms resulting in visually identical estimates.
Figure 2: Slices of the log-likelihood function for $\mu$ (left) and $c$ (right) for the new algorithm (blue line) versus the particle filter (red dots). The particle sizes are $N = 10000$ (top row), $N = 30000$ (middle row), and $N = 100000$ (bottom row). The vertical (green) lines are the ML estimates of the model reported in Table 1 of the paper.
Figure 3: Slices of the log-likelihood function for $\beta$ for the new algorithm (blue line) versus the particle filter (red dots). The particle sizes are $N = 10000$ (top row), $N = 30000$ (middle row), and $N = 100000$ (bottom row). The vertical (green) lines are the ML estimates of the model reported in Table 1 of the paper.

References


