Online Supplemental Appendices for

Testing distributional assumptions using a continuum of moments

Dante Amengual
CEMFI
<amengual@cemfi.es>

Marine Carrasco
Université de Montreal
<marine.carrasco@umontreal.ca>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

March 2017
Revised: June 2019

A Computational details

A.1 Theoretical covariance operator

A.1.1 Eigenvalues and eigenfunctions

As we mentioned in Section 2, the eigenvalues and eigenfunctions of the covariance operator K are the solutions to the functional equations

$$(K\phi_j)(s) = \int [\psi_0(t-s) - \psi_0(t)\psi_0(-s)]\phi_j(t)\pi(t)dt = \lambda_j \phi_j(s).$$

Given that it is not possible to find the analytical solution to this equation for arbitrary distributions, we solve for $\phi_j(s)$ at a very fine but discrete grid of m points over a finite range of values of the characteristic function argument t as follows. For the sake of brevity, here we describe the case in which t is scalar. Let $F(.)$ and $Q(.)$ denote the cdf and quantile functions, respectively, associated with the continuous density function $\pi(t)$, which we assume integrates to 1 over (t_l, t_u). Then, if we define $\nu = F(t)$, the usual change of variable formula immediately implies that the integral between t_l and t_u of any function $g(t)$ weighted by $\pi(t)dt$ coincides with the integral between 0 and 1 of $g[Q(\nu)]d\nu$. We exploit this equivalence to numerically approximate all the required integrals using the rectangle method over m equidistant points between 0 and 1 regardless of $\pi(.)$.

Let $\mathcal{K}$ be an $m \times m$ matrix whose elements are

$$\psi[Q(\nu_i) - Q(\nu_j)] - \psi[Q(\nu_i)]\psi[-Q(\nu_j)], \quad i, j = 1, \ldots, m,$$

so that $\mathcal{K}$ effectively gives us the asymptotic covariance matrix of the sample average of an $m \times 1$ vector of influence functions $e^{iQ(\nu_j)x_l} - \psi[Q(\nu_j)]$, $j = 1, \ldots, m$.

Given that the eigenvalues of $\mathcal{K}$ increase with m, we work with $m^{-1}\mathcal{K}$, whose eigenvalues stabilize. In this context, we take the decreasingly ordered eigenvalues of this scaled matrix as an approximation to the decreasingly ordered eigenvalues of the theoretical covariance operator K. Similarly, we also take the normalized eigenvectors of $m^{-1}\mathcal{K}$ multiplied by $\sqrt{m}$ as an approximation to the eigenfunctions of the covariance operator scaled so that they have unit norm.

A.1.2 Test statistic

We compute the (scaled by $\sqrt{n}$) average values of the "population principal components" of the vector of influence functions $e^{iQ(\nu_j)x_l} - \psi[Q(\nu_j)]$, $j = 1, \ldots, m$ by premultiplying the scaled sample average of this vector by the eigenfunctions previously computed and dividing the resulting expression by m.

Finally, we compute the T_B test statistic as a linear combination of the square norm of the scaled average values of those principal components weighted by $\frac{\lambda_j}{\lambda_j^2+\alpha}$. In effect, this T_B is numerically identical to the overidentifying restriction statistic of a discrete GMM procedure based on the $m \times 1$ vector of influence functions $e^{iQ(\nu_j)x_l} - \psi[Q(\nu_j)]$, in which we replace the inverse of the asymptotic covariance matrix $m^{-1}\mathcal{K}$ by its Tikhonov regularized inverse, as in (11).

A.2 Analytical expressions for c_{il}

A.2.1 Univariate normal

Given that the CF of the standard normal is $\psi(t) = e^{-\frac{1}{2}t^2}$, $h_i(t)\overline{h_l(t)}$ has the following four terms

$$e^{it(x_i - x_l)} - e^{-\frac{1}{2}t^2 + ix_i t} - e^{-\frac{1}{2}t^2 - ix_l t} + e^{-t^2}.$$

Using a $\mathcal{N}(0, \omega^2)$ density as weighting function π, we obtain

$$c_{il} = c_1(x_i, x_l) - c_2(x_i) - c_2(-x_l) + c_3$$

where

$$c_1(x_i, x_l) = \int e^{it(x_i - x_l)}\pi(t)dt = e^{-\frac{1}{2}\omega^2(x_i - x_l)^2},$$

$$c_2(x) = \int e^{-\frac{1}{2}t^2 + ixt}\pi(t)dt = \frac{e^{-\frac{\omega^2 x^2}{2(1+\omega^2)}}}{\sqrt{1+\omega^2}},$$

and

$$c_3 = \int e^{-t^2}\pi(t)dt = \frac{1}{\sqrt{1+2\omega^2}}.$$

A.2.2 Standardized uniform

Given that the CF of the standardized uniform is

$$\psi(t) = \frac{i}{2\sqrt{3}t}(e^{-i\sqrt{3}t} - e^{i\sqrt{3}t}),$$

$h_i(t)\overline{h_l(t)}$ has the following four terms

$$e^{it(x_i - x_l)} - \frac{ie^{ix_i t}}{2\sqrt{3}t}(e^{-i\sqrt{3}t} - e^{i\sqrt{3}t}) - \frac{ie^{-ix_l t}}{2\sqrt{3}t}(e^{-i\sqrt{3}t} - e^{i\sqrt{3}t}) - \frac{e^{-2i\sqrt{3}t}(e^{2i\sqrt{3}t} - 1)^2}{12t^2}.$$

Using a $\mathcal{N}(0, \omega^2)$ density as weighting function π, we obtain

$$c_{il} = c_1(x_i, x_l) - c_2(x_i) - c_2(-x_l) + c_3$$

where

$$c_1(x_i, x_l) = \int e^{it(x_i - x_l)}\pi(t)dt = e^{-\frac{1}{2}\omega^2(x_i - x_l)^2},$$

$$\begin{aligned} c_2(x) &= \int \frac{ie^{ixt}}{2\sqrt{3}t}(e^{-i\sqrt{3}t} - e^{i\sqrt{3}t})\pi(t)dt \\ &= \frac{1}{2\omega}\sqrt{\frac{\pi}{6}}\left\{\operatorname{erf}\left[\frac{\omega(\sqrt{3}-x)}{\sqrt{2}}\right] + \operatorname{erf}\left[\frac{\omega(\sqrt{3}+x)}{\sqrt{2}}\right]\right\}, \end{aligned}$$

where erf is the error function i.e. $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$, and

$$c_3 = -\int \frac{e^{-2i\sqrt{3}t}(e^{2i\sqrt{3}t}-1)^2}{12t^2}\pi(t)dt = \frac{e^{-6\omega^2} - 1 + \sqrt{6\pi}\omega \operatorname{erf}(\sqrt{6}\omega)}{6\omega^2}.$$

A.2.3 Bivariate standard normal

Given that the CF of the bivariate normal with zero mean and identity covariance matrix is $\psi(t_1, t_2) = e^{-\frac{1}{2}(t_1^2+t_2^2)}$, $h_i(t_1, t_2)\overline{h_l(t_1, t_2)}$ has the following four terms

$$e^{i[t_1(x_{1i}-x_{1l})+t_2(x_{2i}-x_{2l})]} - e^{-\frac{1}{2}(t_1^2+t_2^2)+i(t_1x_{1i}+t_2x_{i2})} - e^{-\frac{1}{2}(t_1^2+t_2^2)-i(t_1x_{1i}+t_2x_{i2})} + e^{-t_1^2-t_2^2}.$$

Using two independent $\mathcal{N}(0, \omega^2)$ densities as weighting functions π for both t_1 and t_2, we obtain

$$c_{il} = c_1(x_i, x_l) - c_2(x_i) - c_2(-x_l) + c_3$$

where

$$c_1(x_i, x_l) = \int\int e^{i[t_1(x_{1i}-x_{1l})+t_2(x_{2i}-x_{2l})]}\pi(t_1)\pi(t_2)dt_1dt_2 = e^{-\frac{1}{2}\omega^2[(x_{1i}-x_{1l})^2+(x_{2i}-x_{2l})^2]},$$

$$c_2(x) = \int\int e^{-\frac{1}{2}(t_1^2+t_2^2)+i(t_1x_1+t_2x_2)}\pi(t_1)\pi(t_2)dt_1dt_2 = \frac{e^{-\frac{\omega^2(x_1^2+x_2^2)}{2(1+\omega^2)}}}{(1+\omega^2)},$$

and

$$c_3 = \int\int e^{-t_1^2-t_2^2}\pi_1(t_1)\pi_2(t_2)dt_1dt_2 = \frac{1}{1+2\omega^2}.$$

A.2.4 Standardized chi-square with 2 degrees of freedom

Given that the CF of the standardized $\chi^2(2)$ is $\psi(t) = ie^{-it}/(i+t)$, $h_i(t)\overline{h_l(t)}$ has the following four terms

$$e^{it(x_i-x_l)} - \frac{ie^{it(1+x_i)}}{i+t} - \frac{ie^{it(1+x_i)}}{i-t} - \frac{ie^{-it(1+x_l)}}{i+t} - \frac{1}{(i-t)(i+t)}.$$

Using a $U(-\omega, \omega)$ density as weighting function π, we obtain

$$c_{il} = c_1(x_i, x_l) - c_2(x_i) - c_2(-x_l) + c_3$$

where

$$c_1(x_i, x_l) = \int e^{it(x_i-x_l)}\pi(t)dt = \frac{\sin[\omega(x_i-x_l)]}{\omega(x_i-x_l)},$$

$$
\begin{aligned}
c_2(x) &= \int \frac{ie^{it(1+x)}}{i+t}\pi(t)dt \\
&= \frac{e^{-(1+x)}}{2\omega}\left\{\pi - i\,\mathrm{Ci}[(\omega-i)(1+x)] + i\,\mathrm{Ci}[(\omega+i)(1+x)]\right\} \\
&\quad + \frac{e^{-(1+x)}}{2\omega}\left\{\mathrm{Si}[(\omega+i)(1+x)] - \mathrm{Si}[(i-\omega)(1+x)]\right\},
\end{aligned}
$$

where Si is the sine integral function $\mathrm{Si}(z) = \int_0^z \frac{\sin(t)}{t}dt$, Ci is the cosine integral function $\mathrm{Ci}(z) = -\int_z^\infty \frac{\cos(t)}{t}dt$, and

$$
c_3 = -\int \frac{\pi(t)dt}{(i-t)(i+t)} = \frac{\arctan(\omega)}{\omega}.
$$

A.2.5 Univariate Cauchy

Given that the CF of the univariate Cauchy with location and scale parameters μ and γ, respectively, is $\psi(t) = e^{\mu it+\gamma|t|}$, $h_i(t)\overline{h_l(t)}$ has the following four terms

$$
e^{it(x_i-x_l)} - e^{-\mu it-\gamma|t|+ix_it} - e^{\mu it-\gamma|t|-ix_lt} + e^{-2\gamma|t|}.
$$

Using a $\mathcal{N}(0,\omega^2)$ density as weighting function π, we obtain

$$
c_{il} = c_1(x_i,x_l) - c_2(x_i) - c_2(-x_l) + c_3
$$

where

$$
c_1(x_i,x_l) = \int e^{it(x_i-x_l)}\pi(t)dt = e^{-\frac{1}{2}\omega^2(x_i-x_l)^2},
$$

$$
\begin{aligned}
c_2(x) &= \int e^{-\mu it-\gamma|t|+ix_it}\pi(t)dt \\
&= \frac{1}{2}\left\{e^{\frac{\omega^2}{2}[\gamma-i(x_i-\mu)]^2}\,\mathrm{erf\,c}\left[\frac{\gamma-i(x_i-\mu)}{\sqrt{2}}\right] + e^{\frac{\omega^2}{2}[\gamma+i(x_i-\mu)]^2}\,\mathrm{erf\,c}\left[\frac{\gamma+i(x_i-\mu)}{\sqrt{2}}\right]\right\},
\end{aligned}
$$

and

$$
c_3 = \int e^{-t^2}\pi(t)dt = e^{2\gamma^2\omega^2}\,\mathrm{erf\,c}(\sqrt{2}\gamma\omega).
$$

A.3 Classical goodness of fit tests

We briefly review below some classical goodness of fit tests (see for instance, Lehmann and Romano (2005)), which serve as benchmarks in our Monte Carlo exercise. For convenience, we present them for scalar X.

For testing $H_0 : F = F_0$ versus $H_1 : F \neq F_0$, the classical Kolmogorov-Smirnov (KS) test is based on a sup norm of the difference between the empirical distribution function $\hat{F}_n$ and the distribution function:

$$
KS = \sup_{x\in\mathbb{R}} \sqrt{n}\left|\hat{F}_n(x) - F_0(x)\right|.
$$

On the other hand, the Cramer-von-Mises (CvM) test is based on the L^2 norm of the difference:

$$CvM = n\int_{-\infty}^{\infty}[\hat{F}_n(x) - F_0(x)]^2 dF_0(x).$$

Finally, the Anderson-Darling (AD) test differs from the Cramer-von-Mises by the weight:

$$AD = n\int_{-\infty}^{\infty}\frac{[\hat{F}_n(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} dF_0(x).$$

So far, F_0 was completely specified. For testing normality with unknown mean and variance, the KS test is usually computed as

$$KS = \sup_{x\in\mathbb{R}}\sqrt{n}\left|\hat{F}_n(x) - \Phi\left(\frac{x-\bar{X}}{\hat{\sigma}}\right)\right|$$

where Φ is the distribution function of the standard normal and $\bar{X}$ and $\hat{\sigma}^2$ are the maximum likelihood estimators of the mean and variance. This version of the KS test is often referred to as the Lilliefors test. The other tests can be similarly modified. A multivariate extension is proposed in Andrews (1997).

Consider now the case $X_j \in \mathbb{R}^q$. To test $H_0 : \psi = \psi_0(.;\theta_0)$ versus $H_1 : \psi \neq \psi_0(.;\theta_0)$, Bierens and Wang (2012) consider a L^2 test based on the empirical characteristic function and a uniform weight:

$$BW = \int_{\Upsilon}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}[e^{i\tau' X_j} - \psi_0(.;\hat{\theta})]\right|^2 \frac{d\tau}{2^q\Pi_{l=1}^q\overline{\tau}_l}$$

where $\Upsilon = \times_{l=1}^q[-\overline{\tau}_l, \overline{\tau}_l]$, $\overline{\tau}_l > 0$ and $\hat{\theta}$ is a consistent estimator of θ.

These four tests are consistent against any fixed alternative to the null hypothesis and have power against $1/\sqrt{n}$ alternatives too. However, for testing general distributions with unknown parameter, their asymptotic distributions are not nuisance parameter free.

A.4 On simulating distributions

We simulate all the distributions under the null, as well as the symmetric Student t, gamma and beta distributions, using the available MATLAB routines. Namely, we use `rand.m` for the uniform, `randn.m` (`mvnrnd.m`) for the univariate (bivariate) normal, `chi2rnd.m` for the $\chi^2(2)$, `trnd.m` (`mvtrnd.m`) times $\sqrt{(\nu-2)/2}$ where ν denotes the degrees of freedom for the univariate (bivariate) symmetric Student t, `gamrnd.m` for the gamma and `betarnd.m` for the beta distribution. As for the remaining ones, the procedure is as follows.

A.4.1 Asymmetric Student *t*

The asymmetric t distribution is a special case of the Generalized Hyperbolic family with $\gamma = 0$ and $-\infty < \nu < -2$ (see Mencía and Sentana (2012)). As explained by these authors, if the number of degrees of freedom exceeds 4, we can easily simulate a standardized (zero mean, unit variance) version of a univariate asymmetric Student t distribution by exploiting its representation as a location-scale mixture of normals,

$$X_i = c(\beta,\nu,\gamma)\beta\left[\frac{1-2\eta}{\eta\xi_t}-1\right]+\sqrt{\frac{1-2\eta}{\eta\xi_i}}\sqrt{c(\beta,\nu,\gamma)}Z_i, \tag{A1}$$
$$c(\beta,\nu,\gamma) = \frac{1-4\eta}{2\eta}\frac{\sqrt{1+8\beta'\beta\eta/(1-4\eta)}-1}{2\beta'\beta}$$

where $\eta = -1/(2\nu)$, ξ_i is distributed *iid* gamma with mean η^{-1} and variance $2\eta^{-1}$, and $Z_i|\xi_i$ is *iid* $\mathcal{N}(0,1)$.

If we further assume that $\eta < 1/8$, then the skewness and kurtosis coefficients of the asymmetric t distribution will be

$$E(X_i^3) = 16c^3(\beta,\nu,\gamma)\frac{\eta^2}{(1-4\eta)(1-6\eta)}\beta^3 + 6c^2(\beta,\nu,\gamma)\frac{\eta}{1-4\eta}\beta$$

and

$$E(X_i^4) = 12c^4(\beta,\nu,\gamma)\frac{\eta^2(10\eta+1)}{(1-4\eta)(1-6\eta)(1-8\eta)}\beta^4$$
$$+12c^3(\beta,\nu,\gamma)\frac{\eta(2\eta+1)}{(1-4\eta)(1-6\eta)}\beta^2 + 3\frac{1-2\eta}{1-4\eta}c^2(\beta,\nu,\gamma).$$

Not surprisingly, we can obtain maximum asymmetry for a given kurtosis by letting $|\beta| \to \infty$.

In contrast, a standardized version of the usual symmetric Student t with $1/\eta$ degrees of freedom is achieved when $\beta = 0$ for $\eta < 1/2$. Since $\lim_{\beta\to 0} c(\beta,\nu,\gamma) = 1$, in that case the coefficient of kurtosis becomes

$$E(X_i^4) = 3\frac{1-2\eta}{1-4\eta}$$

for any $\eta < 1/4$, while the coefficient of asymmetry is obviously 0.

In the bivariate case the same location scale interpretation in (A1) applies but with $Z_{it}|\xi_i$ being *iid* $\mathcal{N}(0,I)$. However, since the elements of the resulting random vector are correlated when $\beta \neq 0$, we use the standardization procedure in Mencía and Sentana (2012).

We chose 12 degrees of freedom and $\beta = -.75$ to avoid having too much power for both the univariate and bivariate cases. According to the above calculations, in the univariate case $E(X_i^4) = 3.75$ for the symmetric Student t, while for its asymmetric version, $E(X_i^3) = -.54$ and $E(X_i^4) = 4.62$.

A.4.2 Discrete location-scale mixtures of normals

Univariate discrete location-scale mixtures of normals (DLSMN) Let s_i denote an *iid* Bernoulli variate with $P(s_i = 1) = \lambda$. If $z_i|s_i$ is *iid* $N(0,1)$, then

$$X_i = \frac{1}{\sqrt{1+\lambda(1-\lambda)\delta^2}}\left[\delta(s_i-\lambda)+\frac{s_i+(1-s_i)\sqrt{\varkappa}}{\sqrt{\lambda+(1-\lambda)\varkappa}}Z_i\right],$$

where $\delta \in \mathbb{R}$ and $\varkappa > 0$, is a two component mixture of normals whose first two unconditional moments are 0 and 1, respectively. The intuition is as follows. First, note that $\delta(s_t - \lambda)$ is a shifted and scaled Bernoulli random variable with 0 mean and variance $\lambda(1-\lambda)\delta^2$. But since

$$\frac{s_t + (1-s_t)\sqrt{\varkappa}}{\sqrt{\lambda + (1-\lambda)\varkappa}} Z_t$$

is a discrete scale mixture of normals with 0 unconditional mean and unit unconditional variance that is orthogonal to $\delta(s_t - \lambda)$, the sum of the two random variables will have variance $1 + \lambda(1-\lambda)\delta^2$, which explains the scaling factor.

An equivalent way to define and simulate the same standardized random variable is as follows

$$X_i = \begin{cases} N[\mu_1^*(\eta), \sigma_1^{*2}(\eta)] \text{ with probability } \lambda \\ N[\mu_2^*(\eta), \sigma_2^{*2}(\eta)] \text{ with probability } 1-\lambda \end{cases} \tag{A2}$$

where $\eta = (\delta, \varkappa, \lambda)'$ and

$$\begin{aligned}
\mu_1^*(\eta) &= \frac{\delta(1-\lambda)}{\sqrt{1+\lambda(1-\lambda)\delta^2}}, \\
\mu_2^*(\eta) &= -\frac{\delta\lambda}{\sqrt{1+\lambda(1-\lambda)\delta^2}} = -\frac{\lambda}{1-\lambda}\mu_1^*(\eta), \\
\sigma_1^{*2}(\eta) &= \frac{1}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\varkappa]}, \\
\sigma_2^{*2}(\eta) &= \frac{\varkappa}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\varkappa]} = \varkappa\sigma_1^{*2}(\eta).
\end{aligned}$$

Therefore, we can immediately interpret $\varkappa$ as the ratio of the two variances. Similarly, since

$$\delta = \frac{\mu_1^*(\eta) - \mu_1^*(\eta)}{\sqrt{\lambda\sigma_1^{*2}(\eta) + (1-\lambda)\sigma_1^{*2}(\eta)}},$$

we can also interpret δ as the parameter that regulates the distance between the means of the two underlying components.

We can trivially extended this procedure to define and simulate standardized mixtures with three or more components. Specifically, if we replace the normal random variable in the first branch of (A2) by a k-component normal mixture with mean and variance given by $\mu_1^*(\eta)$ and $\sigma_1^{*2}(\eta)$, respectively, then the resulting random variable will be a $(k+1)$-component Gaussian mixture with zero mean and unit variance.

In the case of two-component Gaussian mixtures, the parameters λ, δ and $\varkappa$ determine the higher order moments of X_i through the relationship

$$E(X_i^j) = \lambda E(x_i^j | s_i = 1) + (1-\lambda) E(x_i^j | s_i = 0),$$

where $E(X_i^j|s_i=1)$ can be obtained from the usual normal expressions

$$\begin{aligned}
&E(X_i|s_t=1)=\mu_1^*(\eta)\\
&E(X_i^2|s_t=1)=\mu_1^{*2}(\eta)+\sigma_1^{*2}(\eta)\\
&E(X_i^3|s_t=1)=\mu_1^{*3}(\eta)+3\mu_1^*(\eta)\sigma_1^{*2}(\eta)\\
&E(X_i^4|s_t=1)=\mu_1^{*4}(\eta)+6\mu_1^{*2}(\eta)\sigma_1^{*2}(\eta)+3\sigma_1^{*4}(\eta)\\
&E(X_i^5|s_t=1)=\mu_1^{*5}(\eta)+10\mu_1^{*3}(\eta)\sigma_1^{*2}(\eta)+15\mu_1^*(\eta)\sigma_1^{*4}(\eta)\\
&E(X_i^6|s_t=1)=\mu_1^{*6}(\eta)+15\mu_1^{*4}(\eta)\sigma_1^{*2}(\eta)+45\mu_1^{*2}(\eta)\sigma_1^{*4}(\eta)+15\sigma_1^{*6}(\eta)
\end{aligned}$$

etc. But since $E(X_i)=0$ and $E(X_i^2)=1$ by construction, straightforward algebra shows that the skewness and kurtosis coefficients will be given by

$$E(X_i^3)=\frac{3\delta\lambda(1-\lambda)(1-\varkappa)}{[\lambda+(1-\lambda)\varkappa][1+\lambda(1-\lambda)\delta^2]^{3/2}}+\frac{\delta^3(1-\lambda)\lambda(1-2\lambda)}{[1+\lambda(1-\lambda)\delta^2]^{3/2}}=a(\delta,\kappa,\lambda) \tag{A3}$$

and

$$\begin{aligned}
E(X_i^4) \quad = \quad & \frac{3[\lambda+(1-\lambda)\varkappa^2]}{[\lambda+(1-\lambda)\varkappa]^2[1+\lambda(1-\lambda)\delta^2]^2}+\frac{6\delta^2\lambda(1-\lambda)[(1-\lambda)+\varkappa\lambda]}{[\lambda+(1-\lambda)\varkappa][1+\lambda(1-\lambda)\delta^2]^2}\\
&+\frac{\delta^4\lambda(1-\lambda)[1-3\lambda(1-\lambda)]}{[1+\lambda(1-\lambda)\delta^2]^2}=b(\delta,\kappa,\lambda).
\end{aligned} \tag{A4}$$

Two issues are worth pointing out. First, $a(\delta,\kappa,\lambda)$ is an odd function of δ, which means that δ and $-\delta$ yield the same skewness in absolute value. In this sense, if we set $\delta=0$ then we will obtain a discrete scale mixture of normals, which is always symmetric but leptokurtic.[5] Second, $b(\delta,\kappa,\lambda)$ is an even function of δ, which implies that δ and $-\delta$ give rise to the same kurtosis. For that reason, in what follows we mostly consider the case of $\delta\geq 0$.

For the symmetric alternatives, we calibrate the parameters by matching the kurtosis coefficient to that one of the Student t with 12 degrees of freedom ($E(X_i^4)=3.75$). Since there are two parameters, we arbitrarily set the probability λ to 1/10 for the so-called "outlier case" (Panel C of Table 1) and to 3/4 for the so-called "inlier case" (Panel B of Table 1), delivering values of $\varkappa$ equal to 1/3 and $(15-8\sqrt{3})/11$, respectively.

As for the asymmetric mixture of three normals, we impose the same skewness and kurtosis as the normal, and fix the fifth and sixth moments to -1 and 18 (as a reference, they are 0 and 15, respectively, in the Gaussian case), which together with arbitrary weights of .3, .3, and .4, allow us to fully characterize the corresponding alternative.

Multivariate scale mixture of two normals $X_i=\sqrt{\varsigma_i}U_i$, with U_i being uniform on the unit sphere surface in $\mathbb{R}^N$, is distributed as a two-point discrete mixture of normals (DSMN) if and only if

$$\varsigma_i\equiv X_i'X_i=\frac{s_i+(1-s_i)\varkappa}{\lambda+(1-\lambda)\varkappa}\varsigma_i^o$$

[5]Another way of obtaining discrete normal mixture distributions that are symmetric is by making $\lambda=\frac{1}{2}$ and $\varkappa=1$.

where s_i is an *iid* Bernoulli variate with $P(s_i = 1) = \lambda$, $\varkappa$ is the variance ratio of the two components, which for identification purposes we restrict to be in the range $(0, 1]$ and ς_i^o is an independent $\chi^2(N)$. The DSMN approaches the multivariate normal when $\varkappa \to 1$, $\alpha \to 1$ or $\alpha \to 0$. Near the limit, though, the distributions can be radically different. For instance, given that $\varkappa \in (0, 1]$ when $\alpha \to 0^+$ there are very few observations with very large variance ("outliers case"), while when $\alpha \to 1^-$ the opposite happens, very few observations with very small variance ("inliers case"). As all scale mixtures of normals, the distribution of x_i is leptokurtic.

We calibrate the bivariate outlier distribution (Panel C of Table 3) by following the same steps as in the univariate case.

A.4.3 Standardized second order Hermite expansion of the standard normal

The standardized version of the density in Lemma 7 that we use as alternative to the univariate normal can be written as

$$f(x; a, b) = \frac{e^{-\frac{1}{2}(a+cx)^2} c[1 + a^2 + acx]}{\sqrt{2\pi}} + \frac{e^{-\frac{1}{2}(a+cx)^2} bc[(a + cx)^2 - 1]}{2\sqrt{\pi}}$$

where $0 < b < \sqrt{2}$ and $a^2 < 2b(\sqrt{2} - b)$ to guarantee the positivity of the density, and with $c = \sqrt{1 - a^2 + b\sqrt{2}}$ to ensure $V(x) = 1$. Moreover, we can obtain an analytical expression for the corresponding cdf in terms of the error function erf,

$$F(x; a, b) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{a + cx}{\sqrt{2}}\right)\right] - \frac{e^{-\frac{1}{2}(a+cx)^2}[a(b + \sqrt{2}) + bcx]}{2\sqrt{\pi}},$$

which is the basis for simulating from this distribution. Specifically, we generate a uniform random number u between 0 and 1 and then numerically find the root x to the equation $F(x; a, b) = u$.

A.4.4 Scaled *F*

If we assume that X_i is *iid* as a standardized symmetric multivariate t with ν degrees of freedom, then

$$X_i = \sqrt{\frac{(\nu - 2)\zeta_i}{\xi_i}} U_i$$

where U_i is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, ζ_i is a $\chi^2(N)$, ξ_i is a $\chi^2(\nu)$, and u_i, ζ_i, and ξ_i are mutually independent. Therefore, we can easily generate a scaled F random variable with mean N from the square Euclidean norm of an N-variate Student t with finite degrees of freedom.

A.4.5 Cauchy

If we assume that X_i is *iid* as a Cauchy with location and scale parameters μ and σ, respectively, then X_i is a non-standardized Student t with location μ, scale σ and one degree of freedom. In particular when $\mu = 0$ and $\sigma = 1$, it is faster to sample it as $X_i = Z_{1i}/Z_{21}$ where Z_{ji} is *iid* –across i and j– standard normal.

A.4.6 Laplace

If we assume that X_i is *iid* as a Laplace with location and scale parameters μ and σ, respectively, then

$$X_i = \mu - b\text{sign}(U_i - \frac{1}{2}) \log(2 - 2|U_i|)$$

where U_i is *iid* uniformly distributed on $[0, 1]$.

B Additional figures

Figure B1: Examples of characteristic functions

Figure B1a: Standard normal

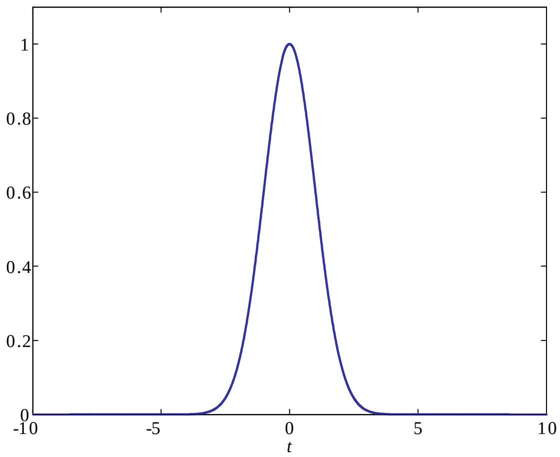


Figure B1b: Standardized uniform

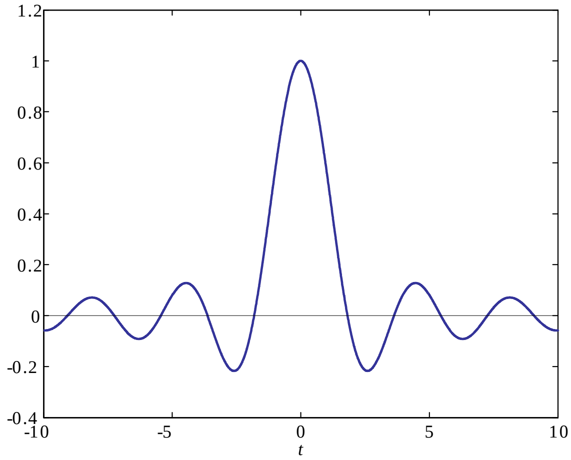


Figure B1c: Standardized $\chi^2(2)$

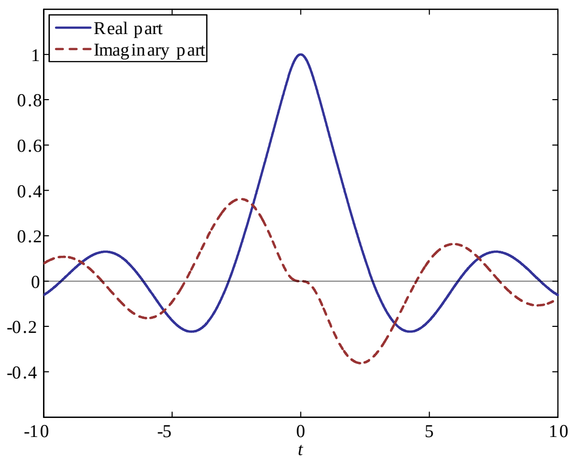


Figure B1d: Standardized Cauchy

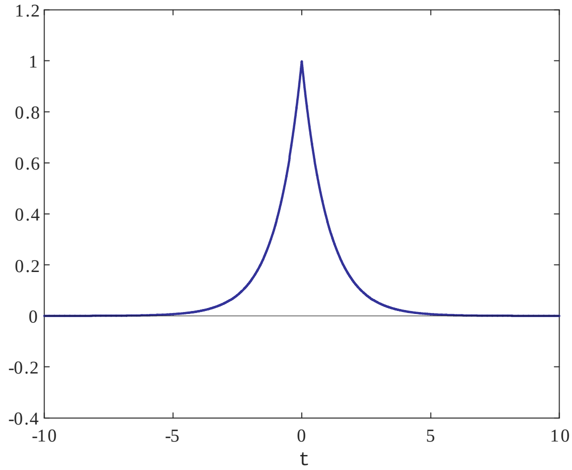


Figure B2: Eigenvalues and eigenfunctions of the covariance operator K

Figure B2a: 1^{st} eigenfunction of K for the standard normal

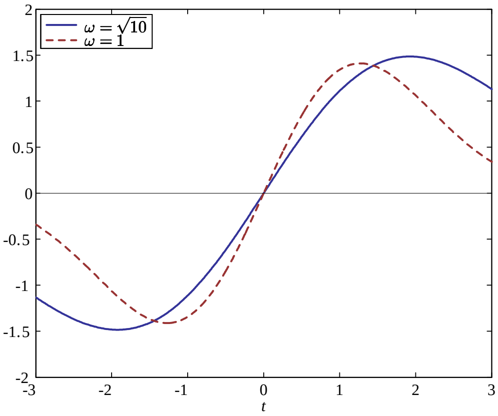


Figure B2b: 1^{st} eigenfunction of K for the (standardized) uniform

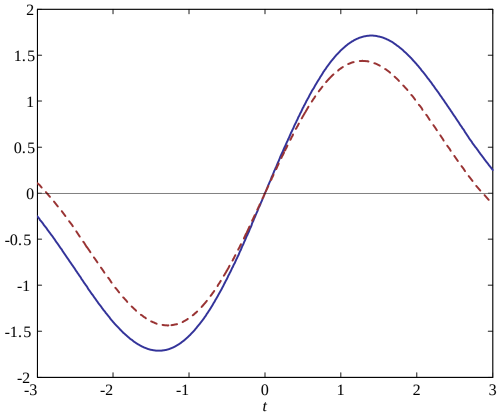


Figure B2c: 2^{nd} eigenfunction of K for the standard normal

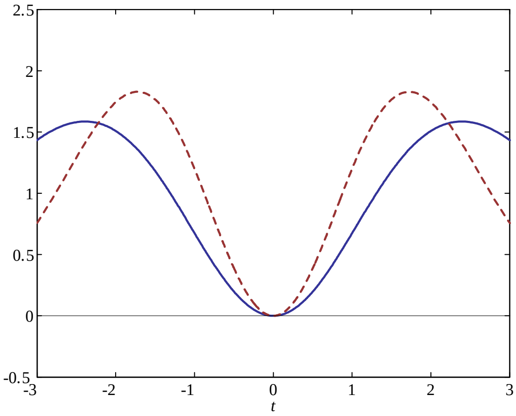


Figure B2d: 2^{nd} eigenfunction of K for the (standardized) uniform

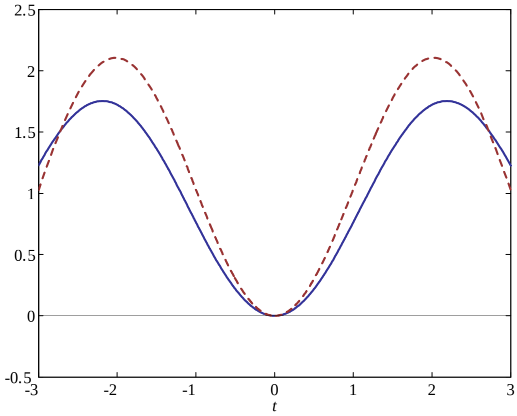


Figure B2e: Eigenvalues of K (in logs) for the standard normal

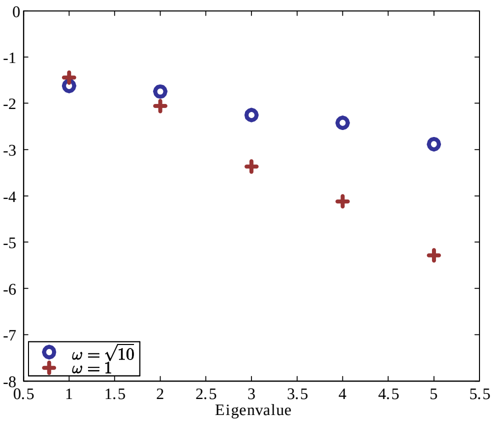
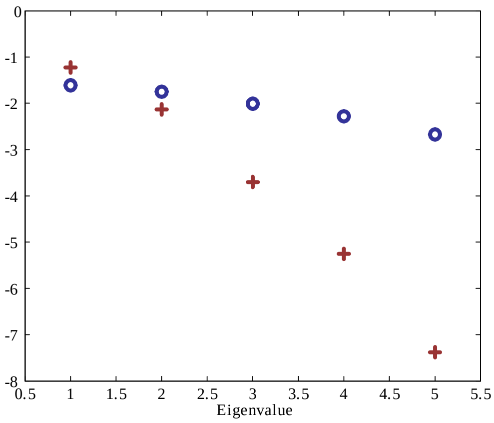


Figure B2f: Eigenvalues of K (in logs) for the (standardized) uniform



Notes: Eigenvalues and eigenfunctions are computed following the procedure described in Appendix A.1 with a grid of 1,000 points. ω is the scale parameter of the $\mathcal{N}(0, \omega^2)$ density defining inner products.

Figure B3: Densities of alternatives to the univariate normal

Figure B3a: Symmetric Student t

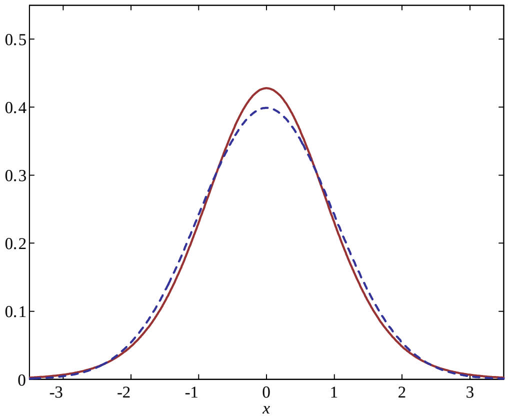


Figure B3b: Asymmetric Student t

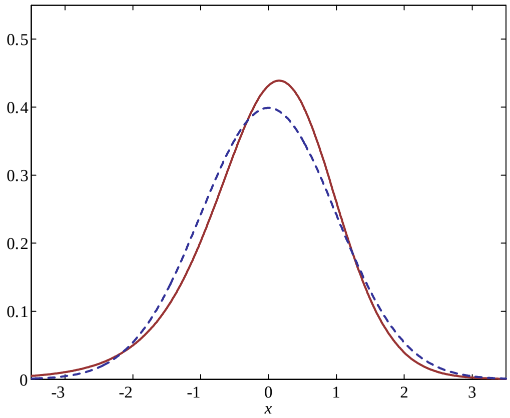


Figure B3c: Scale mixture of two normals (outliers case)

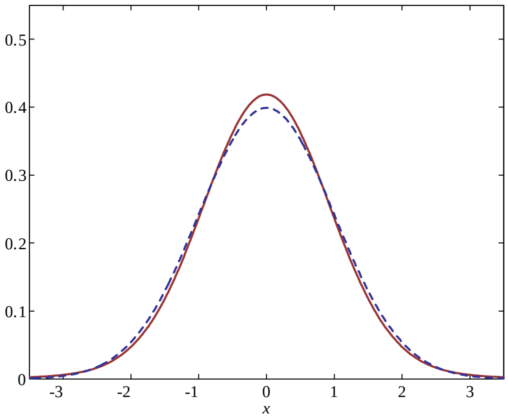


Figure B3d: Third-moment symmetric and mesokurtic Gaussian mixture

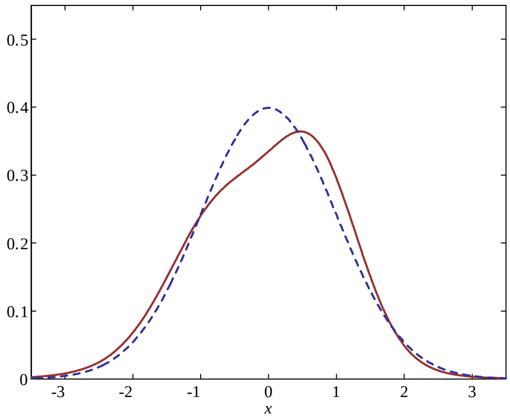


Figure B3e: Scale mixture of two normals (inliers case)

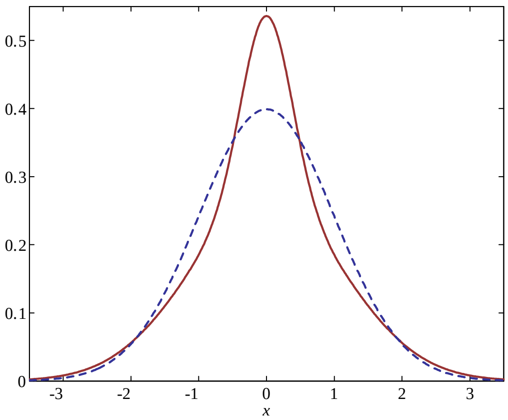
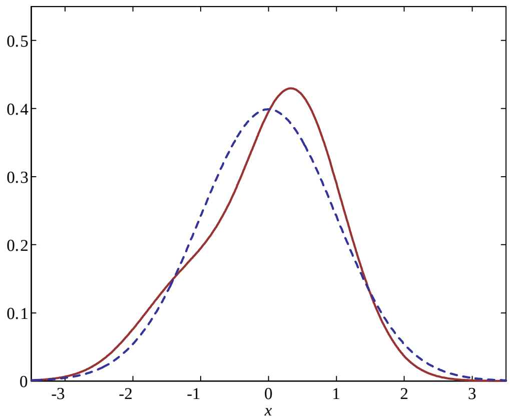


Figure B3f: Second-order Hermite expansion of the normal



Notes: Figure B3a: Student t with 12 degrees of freedom. Figure B3b: Asymmetric t with 12 degrees of freedom and skewness parameter $\beta = -.75$. Figure B3c: Discrete scale mixture with same kurtosis as the symmetric t, 3.75, and $\lambda = 0.1$ (outlier). Figure B3d: Discrete location-scale mixture of three normals with same skewness and kurtosis as the normal and $E(x^5) = -1$, $E(x^6) = 18$. Figure B3e: Discrete scale mixture with kurtosis 3.75 and $\lambda = 0.75$ (inlier). Figure B3f: Second order expansion with $a = 0.4$ and $b = 0.5$. See Appendix A.3 for parameter definitions.

Figure B4: Densities of alternatives to the uniform distribution

Figure B4a: Symmetric beta (with parameters $a = b = 1.1$)

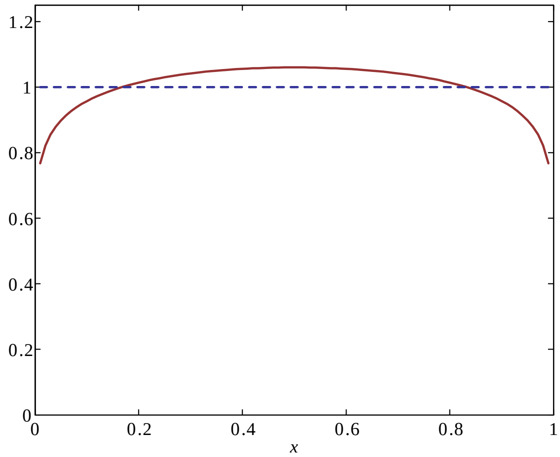


Figure B4b: Asymmetric beta (with parameters $a = 1.1$, $b = 1$)

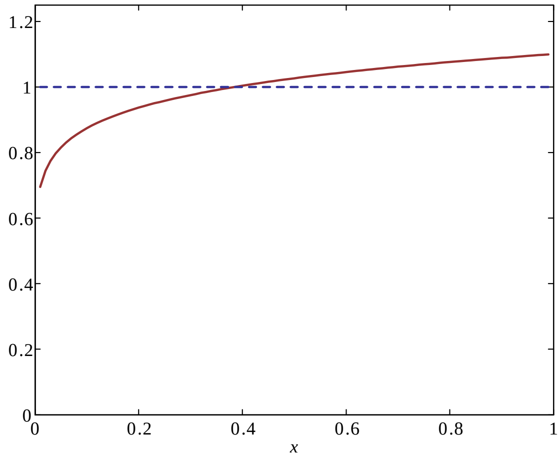
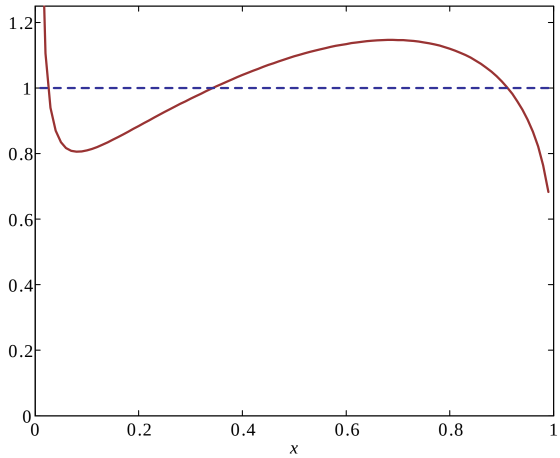


Figure B4c: Gaussian PITs of observations drawn from an asymmetric Student t



Notes: Figure B4c: asymmetric Student t distribution with 12 degrees of freedom and skewness parameter $\beta = -.75$. Density of Figure B4c is computed as the ratio of the pdfs of the asymmetric Student t and the normal.

Figure B5: Alternative distributions to the bivariate normal

Figure B5a: Symmetric Student t density

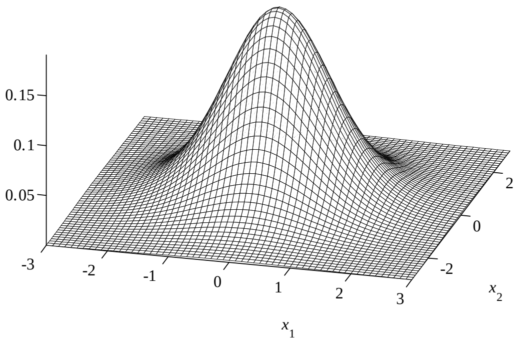


Figure B5b: Contours of a symmetric Student t

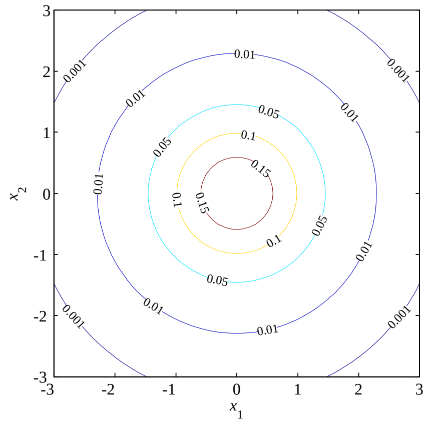


Figure B5c: Scale mixture of two normals (outliers case) density

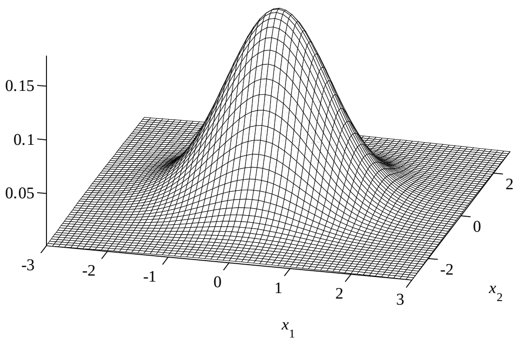


Figure B5d: Contours of a scale mixture of two normals (outliers case)

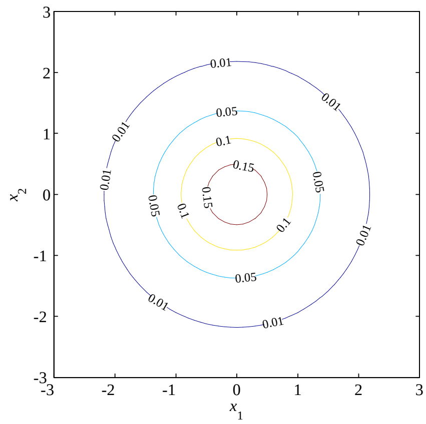


Figure B5e: Asymmetric Student t density

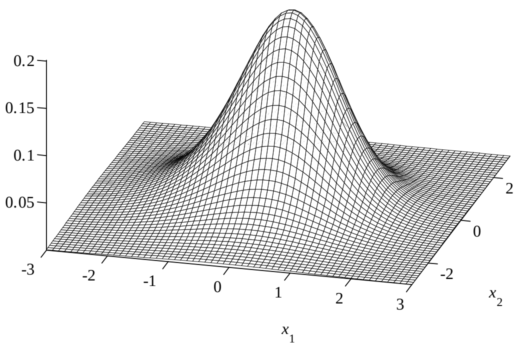
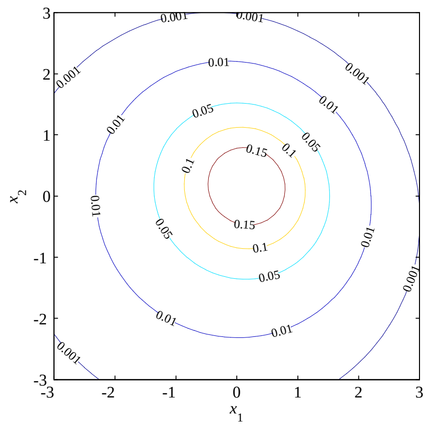


Figure B5f: Contours of an Asymmetric Student t



Notes: Figures B5a–b: Student t with 12 degrees of freedom. Figures B5c–d: Scale mixture with same Mardia's excess kurtosis coefficient as the symmetric t, 0.5, and $\lambda = 0.1$. Figures B5e–f: Asymmetric t with 12 degrees of freedom and skewness parameter $\beta = -.75\ell$. See Appendix B.3 for parameter definitions.

Figure B6: Densities of alternatives to the $\chi^2(2)$

Figure B6a: Scaled F with 2 and 12 degrees of freedom

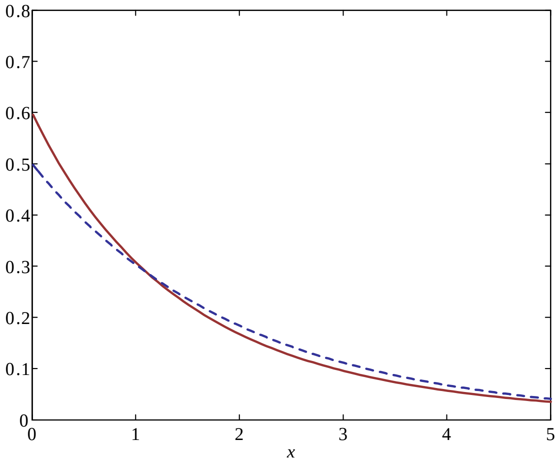


Figure B6b: Gamma with parameters $\alpha = 2/3$ and $\beta = 3$

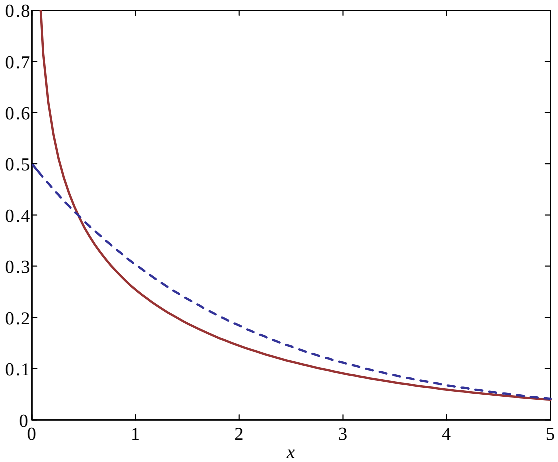
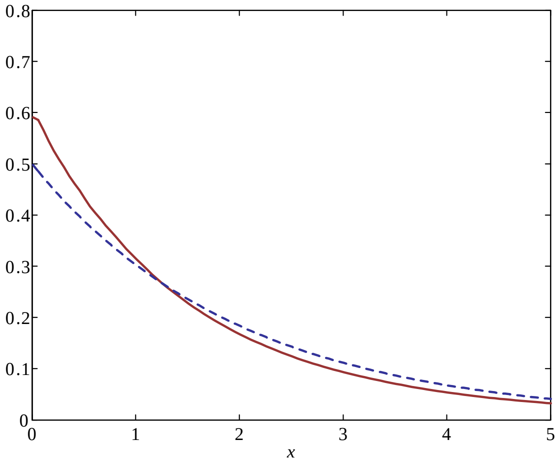


Figure B6c: Square norm of bivariate draws from asymmetric Student t



Notes: Figure B6c: asymmetric Student t distribution with 12 degrees of freedom and skewness parameter vector $\beta = -.75\ell$. Density of Figure B6c was computed by nonparametric estimation of a simulated sample of size 5,000,000.

Figure B7: Densities of alternatives to the Cauchy

Figure B7a: Student t with 2 degrees of freedom

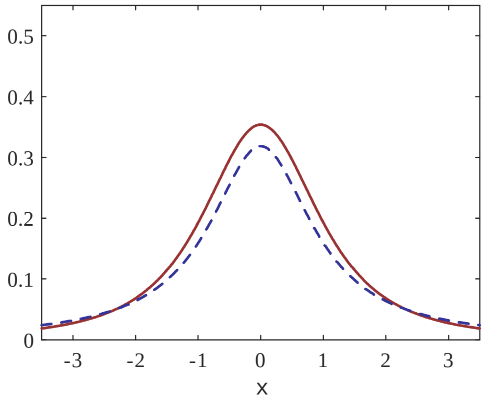


Figure B7b: Asymmetric Student t

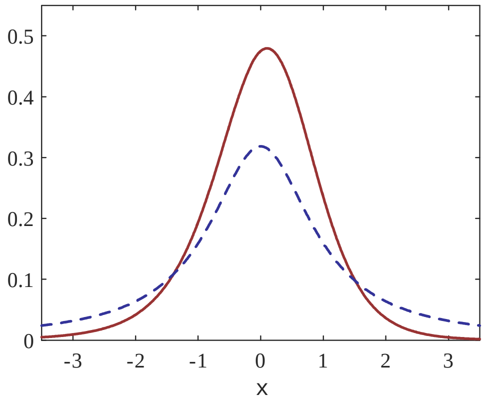
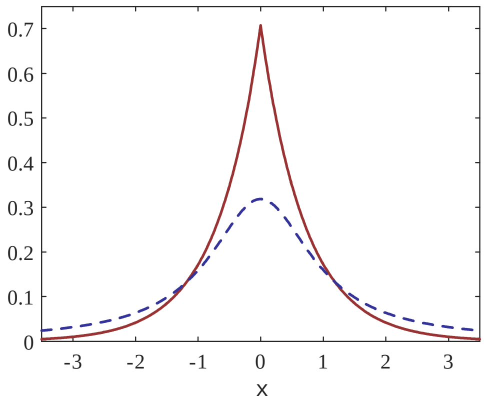


Figure B7c: Laplace with location 0 and scale $1/\sqrt{2}$



Notes: Figure B7a: Student t with 2 degrees of freedom. Figure B7b: Asymmetric Student t with 6 degrees of freedom and skewness parameter $\beta = -.25$. Figure B7c: Laplace with location 0 and scale $1/\sqrt{2}$.