The Multivariate Pearson IV Distribution

by

Mitch Murphy

B.Sc. (Honours), Mathematics, Carleton University (1996)

M.B.A., Sprott School of Business, Carleton University (2006)

Thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy
in
Applied Mathematics

Carleton University
Ottawa, Ontario, Canada

Abstract

The proposed **Multivariate Pearson IV Distribution**, a location-scale family, is defined as

$$f(\vec{\boldsymbol{x}} \mid \vec{\boldsymbol{\mu}}, \boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu) = \frac{\mathcal{C}_N(\boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu)\ e^{\sum_{i=1}^{N} \gamma_i \tan^{-1}\left(\frac{x_i - \mu_i}{\sigma_i \sqrt{\nu}}\right)}}{\left(1 + \frac{(\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}})^T \boldsymbol{\Delta}^{-1} (\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}})}{\nu}\right)^{\frac{\nu+N}{2}}}$$

where $\vec{\boldsymbol{x}} \in \mathbb{R}^N, \vec{\boldsymbol{\mu}} \in \mathbb{R}^N, \boldsymbol{\Delta} \in \mathbb{R}^{N\times N}, \vec{\boldsymbol{\gamma}} \in \mathbb{R}^N, \nu \geq 1,$ and $\mathcal{C}_N(\boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu)$ is denoted the **Pearson IV special function** and investigated from analytical and numerical approaches.

My Ph.D. Thesis is written in a monograph style with (i) an **introduction** motivating the need for innovation in this applied field and a survey of the literature; followed by (ii) an axiomatic **theory** of this univariate, bivariate, and $N-$variate distribution family which can fit most empical datasets with arbitrary mean, variance, skewness, and kurtosis degrees of freedom; (iii) a discussion of best practices in **numerics** of these distributions; and (iv) a survey of current and potential **applications** of these Pearson IV distributions.

Acknowledgements

To my wife, Lorasita Murphy-Nué,

and my parents, Raoul & Liette Murphy.

Examination Board

- **Alan Genz** - Professor, Department of Mathematics, Washington State University.
- **Micheal McIntyre** - Associate Professor, Sprott School of Business, Carleton University.
- **David Amundsen** - Associate Professor, School of Mathematics and Statistics, Carleton University.
- **Victor LeBlanc** - Professor, Department of Mathematics and Statistics, University of Ottawa.
- **Minyi Huang** - Assistant Professor, School of Mathematics and Statistics, Carleton University.

Contents

List of Figures

List of Tables

List of Definitions, Lemmas, Theorems, and Examples

"While these surfaces are of far greater generality than the normal surface of frequency, the writer recognises that they form only a very small section of the whole range of frequency surfaces, even if we confine that range to cases in which $\beta_1, \beta_2, \beta'_1, \beta'_2$ - without regard to higher β's - are supposed to be arbitrary for both variables. [...] **Considerable progress would be made were it feasible to discover a surface for which $r, \sigma_1, \sigma_2, \beta_1, \beta_2, \beta'_1, \beta'_2$ were arbitrary."**

- Karl Pearson (1923). On Non-Skew Frequency Surfaces, Biometrika 15, 231-244.

1 Introduction

The bivariate surface

$$f(\,x, y \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2, \nu\,) \;=$$

$$\frac{\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)}{\sigma_1 \sigma_2} \, \frac{e^{\gamma_1 \tan^{-1}\left(\frac{x-\mu_1}{\sigma_1\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y-\mu_2}{\sigma_2\sqrt{\nu}}\right)}}{\left(1 + \frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{\nu\,(1-\rho^2)}\right)^{\frac{\nu+2}{2}}} \tag{1}$$

is the special case for $N = 2$ of the subject of this Ph.D. thesis, ie. a proposed multivariate distribution which satisfies, for any dimension, the above challenge originally posed by Karl Pearson in the bivariate case almost a century ago (with the exception that both β_2 and β'_2 are represented by the same kurtosis parameter ν)*.

*to completely solve bivariate Pearson (1923) challenge, replace $t_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \nu)$ in (1) with Fang et al. (2002)'s meta-elliptical distribution which has one overall kurtosis degree of freedom, and individual marginal kurtosis degrees of freedom for each dimension.

The Pearson (1895) system of distributions was originally devised to model visibly skewed observations. It was well known at the time how to adjust a theoretical model to fit the first two moments of observed data: Any probability distribution can be extended to a location-scale family. Except in pathological cases, a location-scale family can be made to fit the empirical mean (1^{st} moment) and variance (2^{nd} moment) arbitrarily well. However, it was not known how to construct probability distributions in which the skewness (3^{rd} moment) and kurtosis (4^{th} moment) could be adjusted equally freely. The *Univariate Pearson Distribution System* (Figure 1 and Table 1) solves the problem of empirically fitting univariate skewness and kurtosis.

Nagahara (1999, 2003, 2004); Willink (2008) studied the properties of

$$f(\,x \,|\, \lambda, \alpha, \nu, m\,) = k\left[1 + \left(\frac{x-\lambda}{\alpha}\right)^2\right]^{-m} \exp\left[-\nu \tan^{-1}\left(\frac{x-\lambda}{\alpha}\right)\right] \tag{2}$$

where $m > 1/2$, but the distribution (2) does not converge to a normal distribution as $m \to \infty$ (Figure 2), ie.

$$\lim_{m\to\infty} \left(1+x^2\right)^{-m} = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

while my "corrected" version of the univariate Pearson IV

$$f(\,x \mid \mu, \sigma, \gamma, \nu\,) \;=\; \mathcal{C}(\mu, \sigma, \gamma, \nu) \left(1 + \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} e^{\,\gamma \tan^{-1}\left(\frac{x-\mu}{\sigma\sqrt{\nu}}\right)}$$

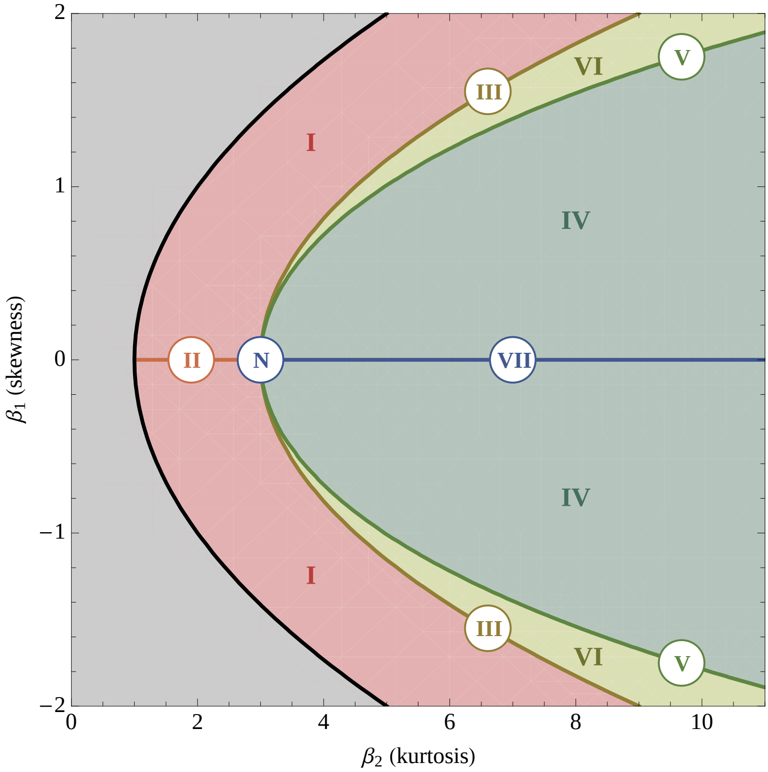


Figure 1: Univariate Pearson Skewness/Kurtosis Diagram

Type	Support	Distribution(s)
Normal Distribution (N)		
		limit of types I, II, III, IV, V, VI and VII
Main Types		
I	(a, b)	Beta
IV	$(-\infty, \infty)$	Pearson IV
VI	(a, ∞)	F, Beta Prime
Transitional Types		
II	$(-a, a)$	*symmetric type I*, eg. Uniform
III	$(-a, \infty)$	χ^2, Gamma, Exponential
V	$(0, \infty)$	Inverse χ^2, Inverse Gamma
VII	$(-\infty, \infty)$	*symmetric type IV*, eg. Student t, Cauchy

Table 1: Univariate Pearson Distribution System (Ord, 1972)

properly converges to a normal distribution as $\nu \to \infty$ (Figure 3), ie.

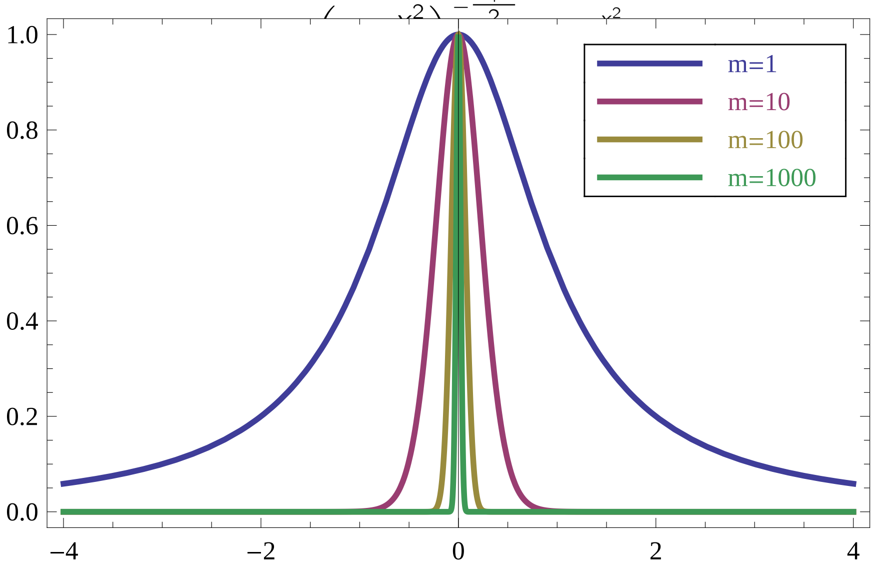


Figure 2: Nagahara's Pearson IV convergence

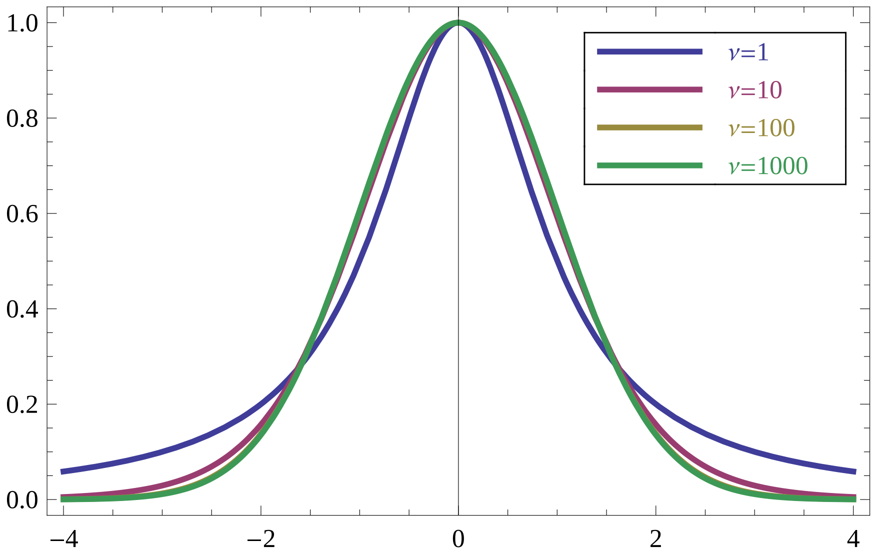


Figure 3: Pearson IV convergence to normal distribution

Kotz and Vicari (2005) provide a detailed historical account of the early 20^{th} century efforts in the area of bivariate distributions by Rhodes (1923, 1925); Van Uven (1925, 1926, 1929, 1947, 1948) and many others. By making specific assumptions concerning the regression and scedastic functions, and requiring that the conditional skewness and kurtosis functions be constant, Narumi (1923) derived the functional forms of a number of bivariate distributions, including the bivariate beta and Student's t families.

Azzalini and Valle (1996) introduced a *skewed multivariate normal distribution* with joint pdf given by

$$f_{\vec{\mathbf{X}}}(\,\vec{\boldsymbol{x}}\,|\,\vec{\boldsymbol{\alpha}},\boldsymbol{\Delta}\,) = 2\;\phi_N(\,\vec{\boldsymbol{x}}\,|\,\boldsymbol{\Delta}\,)\;\Phi(\vec{\boldsymbol{\alpha}}^T\vec{\boldsymbol{x}}), \qquad \vec{\boldsymbol{x}} \in \mathbb{R}^N \tag{3}$$

where $\vec{\boldsymbol{\alpha}} \in \mathbb{R}^N$, $\phi_N(\vec{\boldsymbol{x}};\boldsymbol{\Delta})$ is the multivariate normal density and $\Phi(\cdot)$ is the cdf of the standard univariate normal distribution. Gupta (2003) defined a *skew-t distribution* in a similar fashion to (3) as

$$f_{\vec{\mathbf{X}}}(\,\vec{\boldsymbol{x}}\,|\,\vec{\boldsymbol{\alpha}},\boldsymbol{\Delta},\nu\,) = 2\;t_\nu(\,\vec{\boldsymbol{x}}\,|\,\boldsymbol{\Delta}\,)\;T_{\nu+N}\left(\sqrt{\nu+N}\frac{\vec{\boldsymbol{\alpha}}^T\vec{\boldsymbol{x}}}{\sqrt{\nu+\vec{\boldsymbol{x}}^T\boldsymbol{\Delta}^{-1}\vec{\boldsymbol{x}}}}\right), \qquad \vec{\boldsymbol{x}} \in \mathbb{R}^N \tag{4}$$

where t_ν and $T_{\nu+N}$ denote respectively the joint pdf of the multivariate t distribution and the univariate t cdf. Note that (4) involves a numerical integration to compute the $T_{\nu+N}$ cdf, or a special function evaluation, for each x_i of a sample in maximum likelihood fitting making this distribution computationally inefficient in practice.

Sahu et al. (2003), using transformation and conditioning, obtained another skewed multivariate t distribution given by joint pdf

$$f(\vec{\boldsymbol{x}} \mid \mu, \mathbf{R}, \mathbf{D}) = 2^m \; t_{N,\nu}\left(\vec{\boldsymbol{x}} \mid \mu, \mathbf{R} + \mathbf{D}^2\right) \;\times$$

$$T_{N,\nu+N}\left(\sqrt{\frac{\nu + N}{\nu + \vec{\boldsymbol{y}}^T(\mathbf{R}+\mathbf{D}^2)^{-1}\vec{\boldsymbol{y}}}}\left(\mathbf{I} - \mathbf{D}(\mathbf{R}+\mathbf{D}^2)^{-1}\mathbf{D}\right)^{-1/2}\mathbf{D}(\mathbf{R}+\mathbf{D}^2)^{-1}\vec{\boldsymbol{y}}\right) \quad (5)$$

where $\vec{\boldsymbol{y}} = \vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}}$, $\mathbf{D}$ is a diagonal matrix with the skewness parameters $\delta_1, \ldots, \delta_N$, $t_{N,\nu}(\cdot)$ denotes the joint pdf of the multivariate t distribution, and $T_{N,\nu+N}(\cdot)$ denotes the joint cdf of $t_{N,\nu}(\vec{\mathbf{0}}, \mathbf{I})$. Note that (5) involves several matrix and vector operations, making this distribution vulnerable to performance degradation due to the "curse of dimensionality".

Jones (2002), using a *marginal replacement*

$$f_1(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}) = g(\vec{\boldsymbol{x}})\frac{f(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}})}{f_X(\vec{\boldsymbol{x}})},$$

obtained yet another skewed t distribution with joint pdf

$$\frac{2^{1-a-c}\Gamma\left(\frac{\nu+N}{2}\right)\Gamma(a+c)}{(\pi\nu)^{\frac{N-1}{2}}\sqrt{a+c}\,\Gamma\left(\frac{\nu+1}{2}\right)\Gamma(a)\Gamma(c)} \times$$

$$\left(1+\frac{x_1^2}{\nu}\right)^{\frac{\nu+1}{2}}\left(1+\frac{x_1}{\sqrt{a+c+x_1^2}}\right)^{a+\frac{1}{2}}\left(1-\frac{x_1}{\sqrt{a+c+x_1^2}}\right)^{c+\frac{1}{2}}\left(1+\frac{\vec{\boldsymbol{x}}^T\vec{\boldsymbol{x}}}{\nu}\right)^{-\frac{\nu+N}{2}} \quad (6)$$

which has (i) a skewed t marginal (Jones and Faddy, 2003) with parameters a and c

in the x_1 dimension; (ii) bivariate t conditional distributions for $X_i \,|\, X_1$ $(i = 2, \ldots, N)$; and (iii) a trivial correlation structure $\vec{\boldsymbol{\sigma}}^T \mathbf{I}$.

In the context of kinetic gas theory, Torrilhon (2010) proposed a trivariate Pearson IV distribution

$$f(\vec{\boldsymbol{x}}; \vec{\boldsymbol{\lambda}}, \mathbf{A}, \bar{m}, \nu, \vec{\boldsymbol{n}}) = \frac{K(\bar{m}, \nu)}{\det(\mathbf{A})} \frac{e^{-\nu \tan^{-1}\left(\vec{\boldsymbol{n}}^T \mathrm{A}^{-1}(\vec{\boldsymbol{c}} - \vec{\boldsymbol{\lambda}})\right)}}{\left(1 + (\vec{\boldsymbol{c}} - \vec{\boldsymbol{\lambda}})^T \mathbf{A}^{-2} (\vec{\boldsymbol{c}} - \vec{\boldsymbol{\lambda}})\right)^{\bar{m}}} \tag{7}$$

to model particle velocities with globally hyperbolic moment equations, justified by the fact that it reduces to the Maxwellian distribution in equilibrium and allows for skewness and anisotropies in non-equilibrium to model stress and heat flux.

Faced with the lack of a proper multivariate distribution for my intended financial applications, I was inspired to invent a new multivariate distribution which would have Pearson IV univariate distributions as marginals, non-trivial correlation structure, computationally efficient functional form, and reduce to the multivariate t distribution as skewness approaches zero, ie. when $\vec{\boldsymbol{\gamma}} \to \vec{\mathbf{0}}$.

The proposed *Multivariate Pearson IV (MPIV) distribution*, a location-scale family, is defined as

$$f(\,\vec{\boldsymbol{x}} \mid \vec{\boldsymbol{\mu}}, \boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu\,) \;=\; \mathcal{C}_N(\boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu)\, \frac{e^{\sum_{i=1}^N \gamma_i \tan^{-1}\left(\frac{x_i - \mu_i}{\sigma_i \sqrt{\nu}}\right)}}{\left(1 + \dfrac{(\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}})^T \boldsymbol{\Delta}^{-1} (\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}})}{\nu}\right)^{\frac{\nu+N}{2}}} \tag{8}$$

where

$$\vec{\boldsymbol{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N, \qquad \vec{\boldsymbol{\mu}} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix} \in \mathbb{R}^N, \qquad \vec{\boldsymbol{\gamma}} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_N \end{bmatrix} \in \mathbb{R}^N, \qquad \nu \geq 1,$$

and

$$\boldsymbol{\Delta} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1N}\sigma_1\sigma_N \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2N}\sigma_2\sigma_N \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N}\sigma_1\sigma_N & \rho_{2N}\sigma_2\sigma_N & \cdots & \sigma_N^2 \end{bmatrix}$$

A common strategy of applied mathematics is to "wrap" the complexities of the problem under study into a special function which can be further investigated later. Numerical algorithms, and expansions methods, are tools which can approximate the special function at any given point of its domain for some given accuracy.

We need the integration constant, denoted the *Pearson IV special function ($\mathcal{C}$)*

$$\mathcal{C}_N(\boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu) = \frac{1}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \; f(\,\vec{\boldsymbol{x}} \mid \vec{\boldsymbol{0}}, \boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu\,) \; d\vec{\boldsymbol{x}}} \tag{9}$$

which can be computed numerically by making the substitution $\theta_i = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x_i}{\sigma_i\sqrt{\nu}}\right)$ from $\mathbb{R}^N$ to the unit hypercube $[0,1]^N$, yielding a surface $g(\theta_1, \ldots, \theta_N)$ suitable for the

CUHRE deterministic multivariate numerical integration algorithm (Genz and Cools, 2003; Hahn, 2005).

The Pearson IV special function has dimension-reducing properties

$$\mathcal{C}_1(\mu,\sigma,\gamma,\nu)=\frac{\mathcal{C}_1(\gamma,\nu)}{\sigma},\quad \mathcal{C}_2(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\gamma_1,\gamma_2,\nu)=\frac{\mathcal{C}_2(\rho,\gamma_1,\gamma_2,\nu)}{\sigma_1\,\sigma_2},\quad \ldots \tag{10}$$

and symmetry properties

$$\mathcal{C}_2(-\rho,\gamma_1,\gamma_2,\nu)=\mathcal{C}_2(\rho,-\gamma_1,\gamma_2,\nu)=\mathcal{C}_2(\rho,\gamma_1,-\gamma_2,\nu) \tag{11}$$

and by multiplying a multivariate t distribution (Kotz and Nadarajah, 2004) by $e^{\sum \gamma_i \tan^{-1}\left(\frac{x_i-\mu_i}{\sigma_i\sqrt{\nu}}\right)}$, we know the values of $\mathcal{C}_{1,2,\ldots,N}$ for skewness zero

$$\mathcal{C}_1(0,\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\;\Gamma\left(\frac{\nu}{2}\right)},\qquad \mathcal{C}_2(\rho,0,0,\nu)=\frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\pi\nu\;\Gamma\left(\frac{\nu}{2}\right)\sqrt{1-\rho^2}}, \tag{12}$$

$$\mathcal{C}_N(\mathbf{\Delta},\vec{\mathbf{0}},\nu)=\frac{\Gamma\left(\frac{\nu+N}{2}\right)}{(\pi\nu)^{N/2}\;\Gamma\left(\frac{\nu}{2}\right)\sqrt{|\mathbf{\Delta}|}}$$

to validate the correctness of numerical integration results, together with the obvious check that the pdf integrates to one. Tables of the univariate, bivariate, and trivariate Pearson IV special functions given in §3.1 were obtained by univariate (Piessens et al., 1983) and multivariate (Genz and Malik, 1980a,b; Berntsen et al., 1991; Genz and Cools, 2003) numerical integration.

The univariate Pearson IV distribution is related to the normal, t, and Cauchy distributions (Figure 4)[†], and the multivariate Pearson IV distribution is similarly related to the *multivariate normal (MVN)*, *multivariate t (MVT)*, and *multivariate Cauchy (MVC)* distributions (Figure 5).

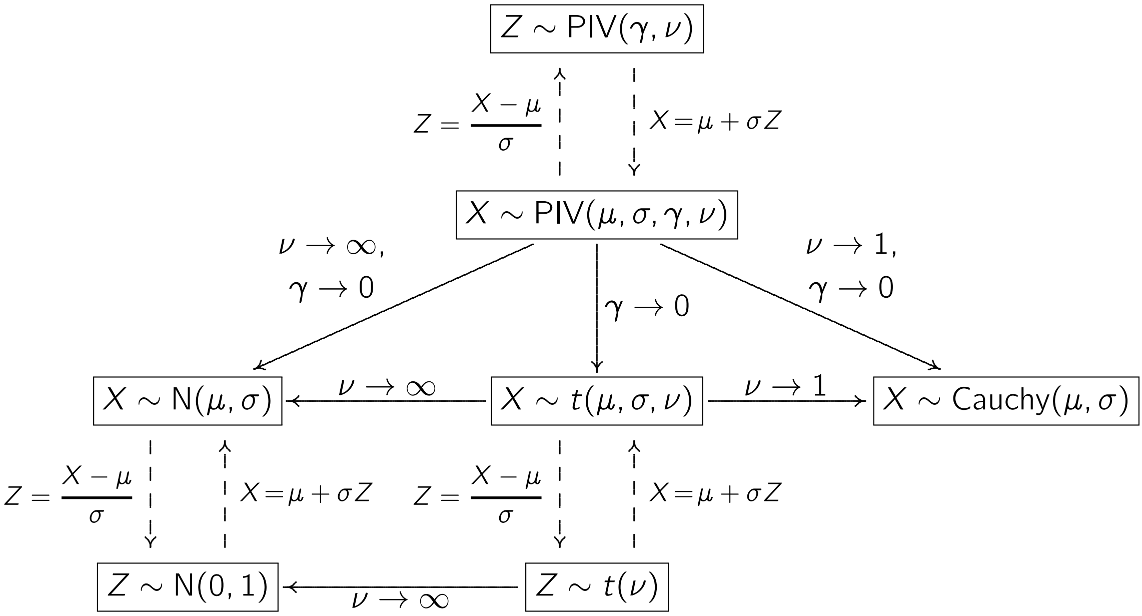


Figure 4: Related Univariate Distributions

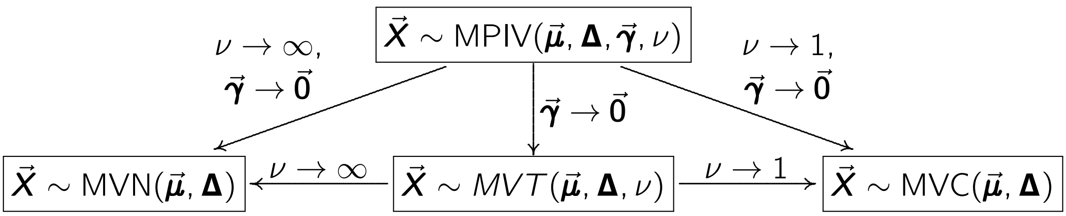


Figure 5: Related Multivariate Distributions

[†]solid lines represent limiting behavior, and dashed lines location-scale transformations.

Note that, unlike multivariate normal (MVN) and multivariate$-t$ (MVT) distributions which are *elliptical distributions*

$$f(x) \propto |\Sigma|^{-1/2} g\left((x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

for some function g, the MPIV distribution is by design intended to capture deviations away from multivariate ellipticality.

Finance is the primary intended application, as there is a need for univariate and multivariate non-normal distributions which can model skewness and kurtosis present in the empirical evidence of many financial assets. Most of the 65 stocks with the largest traded option volumes on the Chicago Board of Options Exchange (CBOE) exhibit non-normal skewness and kurtosis (Table 2). Divergence away from the normal distribution (skewness 0 and kurtosis 3) is most pronounced at small time scales (Figure 6), and the skewness/kurtosis of these time series are located within type IV of the Pearson diagram (Figure 2).

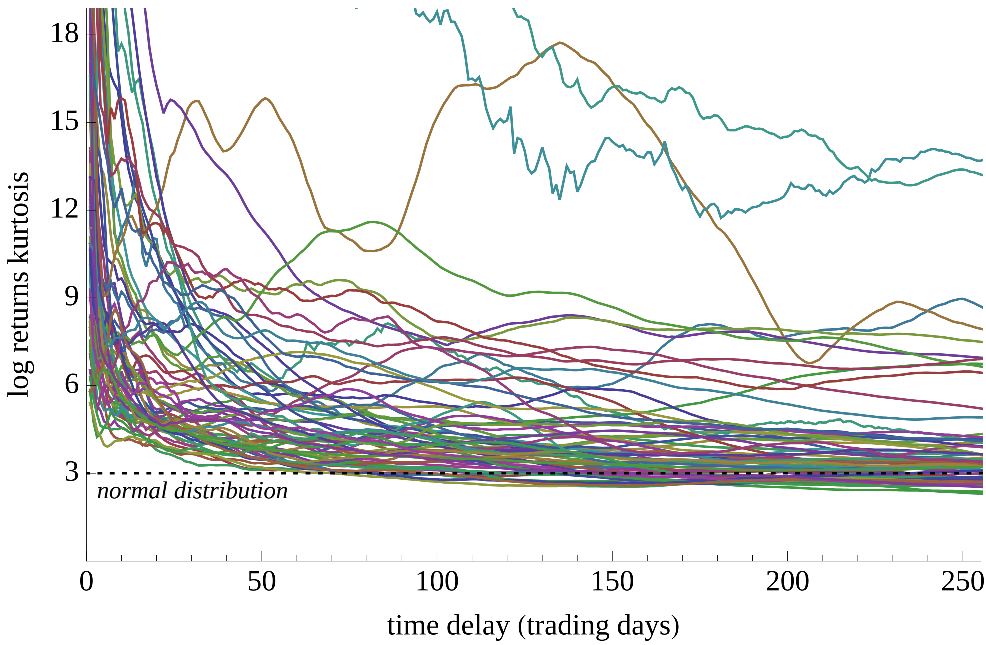


Figure 6: CBOE 65 - Kurtosis vs. time delay

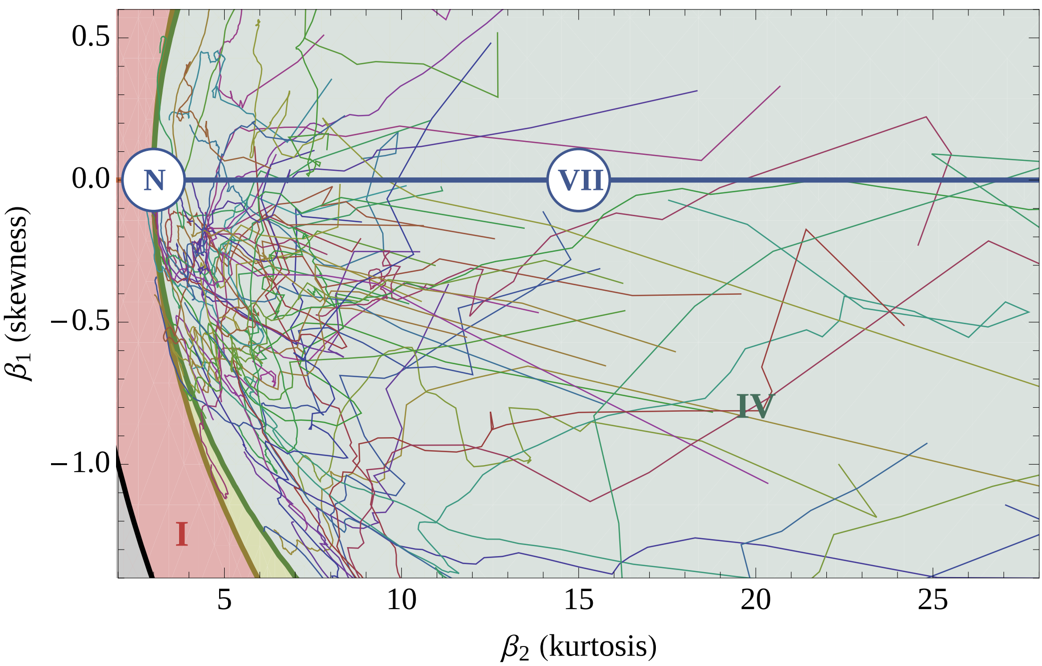


Figure 7: CBOE 65 - Skewness and Kurtosis

	skewness				kurtosis			
	1d	1w	1m	1y	1d	1w	1m	1y
Advanced Micro Devices (AMD)	0.79	0.37	0.39	.49	16.	6.6	4.4	3.1
AK Steel Holding (AKS)	0.37	1.	1.2	0.64	11.	9.6	7.7	3.7
Amazon.Com (AMZN)	.35	.23	.41	.2	8.	5.8	5.	3.1
American International Group (AIG)	8.6	14.	8.9	5.5	280.	360.	120.	42.
Amgen (AMGN)	.075	.11	0.072	.43	8.9	9.4	5.5	3.1
AMR (AMR)	0.23	0.14	0.39	1.1	25.	17.	11.	7.8
Apple (AAPL)	2.	1.3	1.3	0.3	52.	18.	11.	3.2
Applied Materials (AMAT)	.12	0.44	0.48	.29	5.9	6.7	5.5	2.8
AT&T (T)	0.016	0.19	0.41	0.7	8.3	6.1	4.5	3.4
Bank of America (BAC)	0.51	0.81	0.99	0.77	24.	20.	8.8	5.1
Beazer Homes USA (BZH)	0.11	0.68	1.1	1.1	14.	9.9	8.1	5.2
Best Buy (BBY)	1.1	0.33	0.36	.47	20.	7.	4.8	2.9
Boeing (BA)	.1	0.24	0.18	0.18	7.5	6.5	4.6	3.1
Bristol-Myers Squibb (BMY)	0.47	0.64	0.69	1.2	14.	7.4	6.4	6.1
Broadcom (BRCM)	.089	0.19	0.55	0.37	6.5	4.7	4.9	3.
Caterpillar (CAT)	3.6	1.5	1.2	0.2	120.	24.	11.	2.8
Chesapeake Energy (CHK)	0.15	0.35	0.42	0.46	8.9	7.	4.8	3.3
Chevron (CVX)	0.02	0.097	0.24	0.47	10.	5.3	4.6	3.2
Citigroup (C)	.0066	1.5	1.4	0.89	41.	32.	14.	5.1
ConocoPhillips (COP)	0.3	0.28	0.6	0.037	11.	6.7	5.5	3.5
Corning (GLW)	1.	0.88	0.88	1.1	22.	15.	12.	6.5
CVS Caremark (CVS)	0.46	0.63	0.66	0.75	16.	8.5	5.5	4.2
Dell Computer Corp. (DELL)	0.26	0.19	0.32	.21	7.9	5.4	3.9	2.7
eBay (EBAY)	.83	.8	1.7	.87	12.	9.	12.	7.1
EMC (EMC)	0.43	0.3	0.54	0.91	11.	5.3	4.5	4.3
E*Trade Financial (ETFC)	0.97	.17	.26	.17	34.	8.1	6.6	3.5
Exxon Mobil (XOM)	0.4	0.38	0.15	0.21	20.	7.8	4.3	3.1
Ford Motor (F)	0.071	0.46	0.93	0.25	18.	28.	14.	3.8
General Motors (GM)	0.44	1.	1.6	2.1	34.	29.	14.	15.
Halliburton (HAL)	1.3	0.74	1.1	0.91	35.	11.	8.	4.4
Home Depot (HD)	0.65	0.47	0.63	0.26	16.	7.7	5.2	3.2
Intel (INTC)	0.42	0.65	0.73	0.19	9.1	7.1	5.5	3.1
JPMorgan Chase (JPM)	0.36	0.43	0.48	0.4	16.	8.8	5.7	3.6

Table 2: CBOE 65 - Skewness and Kurtosis (1 day, 1 week, 1 month, 1 year)

	skewness				kurtosis			
	1d	1w	1m	1y	1d	1w	1m	1y
McDonald's (MCD)	0.16	0.14	0.47	0.28	11.	5.9	5.8	4.5
Medarex (MEDX)	.11	.08	.32	1.6	7.9	7.1	7.6	7.6
Merck (MRK)	0.82	0.52	0.51	0.27	19.	6.9	4.7	2.9
Micron Technology (MU)	0.075	0.27	0.44	0.033	5.6	5.1	4.7	2.8
Microsoft (MSFT)	0.6	0.3	0.18	.55	18.	7.6	5.	4.1
Motorola (MOT)	0.55	0.59	0.59	0.49	12.	7.1	5.	3.6
NVIDIA (NVDA)	0.21	0.023	0.2	0.57	13.	8.1	4.6	3.6
Oracle (ORCL)	0.17	0.041	0.65	0.17	13.	6.1	5.9	4.3
Pfizer (PFE)	0.26	0.26	0.21	.098	7.3	5.5	4.	2.5
QUALCOMM (QCOM)	.51	.25	.36	1.5	7.8	5.5	4.9	6.3
Rambus (RMBS)	.52	.41	.91	0.48	13.	8.7	7.9	3.6
Research in Motion Limited (RIMM)	0.31	0.66	0.8	0.019	16.	10.	6.8	3.
SanDisk (SNDK)	.21	.032	0.11	.36	11.	6.	3.8	2.8
Schlumberger Limited (SLB)	0.44	0.33	0.63	0.39	10.	5.8	6.	3.
Sirius Satellite Radio (SIRI)	.33	.15	0.05	0.86	21.	8.4	4.8	3.8
SLM (SLM)	0.8	0.44	1.5	1.2	35.	18.	11.	5.8
Sprint Nextel (S)	1.1	1.5	1.5	1.3	27.	22.	10.	4.9
Target (TGT)	0.93	1.3	1.5	0.26	25.	20.	13.	2.9
Texas Instruments Incorporated (TXN)	0.25	0.6	0.39	.1	11.	7.9	4.7	2.9
The Goldman Sachs Group (GS)	.48	0.37	0.95	0.99	13.	8.8	8.3	5.5
Time Warner (TWX)	.041	.16	.24	.27	6.3	4.5	4.	2.6
Transocean (RIG)	0.086	0.28	0.76	0.7	6.4	5.	4.9	2.9
United States Seel (X)	0.21	0.33	1.1	0.21	8.8	6.8	9.1	5.
Valero Energy (VLO)	1.8	.83	.21	0.41	46.	16.	8.2	3.8
Verizon Communications (VZ)	0.024	0.17	0.32	0.37	11.	6.8	4.7	3.4
Wachovia (WB)	18.	9.3	4.7	2.7	930.	290.	68.	15.
Wal-Mart Stores (WMT)	0.012	0.091	0.3	.15	110.	31.	12.	4.8
Washington Mutual (WM)	0.5	0.77	1.3	2.	57.	20.	8.9	10.
Wells Fargo (WFC)	.31	.078	0.098	.2	18.	9.2	5.	3.4
Yahoo! (YHOO)	.23	.21	0.023	0.54	8.4	5.9	4.7	3.7

Table 2: CBOE 65 - Skewness and Kurtosis (1 day, 1 week, 1 month, 1 year)

"A proof is a proof. What kind of a proof? It's a proof. A proof is a proof. And when you have a good proof, it's because it's proven."

- **Jean Chrétien**, Prime Minister of Canada, on the need for evidence of weaponry in Iraq.

2 Theory

2.1 Pearson Differential Equations

Karl Pearson (1895) observed that the normal density function obeys the differential equation

$$\frac{df(x)}{dx} = \frac{(x-a)f(x)}{b}$$

where $a = \mu$, $b = -\sigma^2$, and that the hypergeometric distribution for direct sampling without replacement obeys a difference equation of the form

$$\Delta f_{r-1} \ = \ f_r - f_{r-1} \ = \ \frac{(r-a)f_r}{b_0 + b_1 r + b_2 r^2}, \qquad r \geq 1$$

where f_r is the probability of r successes and a, b_i are parameters. Pearson used these forms to generate his system. Allowing the lattice width to approach zero, we get in the limit the differential equation for density function $f(x)$,

$$\frac{d \ln f(x)}{dx} = \frac{f'(x)}{f(x)} = -\frac{x+a}{b_0 + b_1 x + b_2 x^2} \tag{13}$$

The following lemma is a modified version of a result from Stuart and Ord (1994) and Johnson et al. (2002), with the raw moment μ'_n written as a function of μ'_{n-1} and μ'_{n-2} instead of a relation between μ'_{n+1}, μ'_n, and μ'_{n-1}. This facilitates symbolic implementation using Mathematica, needed for my original result of analytical expressions for the central moments μ_1, μ_2, μ_3, and μ_4 of the univariate Pearson IV distribution (Theorem 2.5).

Lemma 2.1 (moment recurrence relation) *The moments of any Pearson distribution f, assuming $x^n f(x) \to 0$ as $x \to \pm\infty$, follow the recurrence relation*

$$\mu'_n = \frac{(n-1)b_0\mu'_{n-2} + (nb_1 - a)\mu'_{n-1}}{1-(n+1)b_2} \tag{14}$$

with $\mu'_{-1} = 0, \mu_0 = 1$.

Proof.

$$(13) \quad \Rightarrow \quad (b_0 + b_1 x + b_2 x^2) f'(x) = -(x+a) f(x)$$

$$\Rightarrow \quad x^{n-1}(b_0 + b_1 x + b_2 x^2) f'(x) = -x^{n-1}(x+a) f(x)$$

Integrating both sides,

$$\underbrace{x^{n-1}\left(b_0 + b_1 x + b_2 x^2\right) f(x)\Big|_{-\infty}^{\infty}}_{=0,\text{ since } x^n f(x) \to 0 \text{ as } x \to \pm\infty} - \int_{-\infty}^{\infty} \left((n-1)b_0 x^{n-2} + nb_1 x^{n-1} + (n+1)b_2 x^n\right) f(x)dx$$

$$= -\int_{-\infty}^{\infty} x^n f(x)dx - a\int_{-\infty}^{\infty} x^{n-1} f(x)dx$$

$$\Rightarrow \quad -(n-1)b_0\mu'_{n-2} - nb_1\mu'_{n-1} - (n+1)b_2\mu'_n = -\mu'_n - a\mu'_{n-1}$$

$$\Rightarrow \quad \mu'_n = \frac{(n-1)b_0\mu'_{n-2} + (nb_1 - a)\mu'_{n-1}}{1-(n+1)b_2}$$

□

Van Uven (1925, 1926, 1929, 1947, 1948) extended the Pearson differential equation (13) to the bivariate case

$$\frac{\partial \ln f(x,y)}{\partial x} = \frac{A(x,y)}{B(x,y)} = \frac{a_0 + a_1 y + x}{b_0 + b_1 x + b_2 x^2 + b_3 xy + b_4 y + b_5 y^2} \tag{15}$$

$$\frac{\partial \ln f(x,y)}{\partial y} = \frac{C(x,y)}{D(x,y)} = \frac{c_0 + c_1 x + y}{d_0 + d_1 x + d_2 x^2 + d_3 xy + d_4 y + d_5 y^2} \tag{16}$$

with the mixed partial condition

$$\frac{\partial^2 \ln f(x,y)}{\partial x\,\partial y} = \frac{\partial^2 \ln f(x,y)}{\partial y\,\partial x}$$

$$\Rightarrow \quad \frac{\partial}{\partial x}\left(\frac{C(x,y)}{D(x,y)}\right) = \frac{\partial}{\partial y}\left(\frac{A(x,y)}{B(x,y)}\right)$$

$$\Rightarrow \quad \frac{D\dfrac{\partial C}{\partial x} - C\dfrac{\partial D}{\partial x}}{D^2} = \frac{B\dfrac{\partial A}{\partial y} - A\dfrac{\partial B}{\partial y}}{B^2}$$

To solve (15)-(16), van Uven considered all possible forms for $B(x,y)$ and $D(x,y)$. By a complete analysis of all possible cases, Van Uven enumerated a list of bivariate Pearson surfaces (Table 3). Unfortunately, no bivariate surface with marginal Pearson IV distributions was discovered at that time.

Type	Equation	Conditions	Marginal Types	
			x_1	x_2
I	$f(x_1)f(x_2)$	(independent variables with frequencies $f(x_1), f(x_2)$)		
II$a\alpha$	$\dfrac{\Gamma(m_1+m_2+m_3)x_1^{m_1-1}x_2^{m_2-1}(1-x_1-x_2)^{m_3-1}}{\Gamma(m_1)\Gamma(m_2)\Gamma(m_3)}$	$m_1, m_2, m_3 > 0$; $x_1, x_2 > 0$; $x_1 + x_2 \leq 1$	I or II	I or II
II$a\beta$	$\dfrac{\Gamma(1-m_3)x_1^{m_1-1}x_2^{m_2-1}(1+x_1+x_2)^{m_3-1}}{\Gamma(m_1)\Gamma(m_2)\Gamma(1-m_1-m_2-m_3)}$	$m_1, m_2 > 0$; $m_1+m_2+m_3 < 1$; $x_1, x_2 > 0$	VI	VI
II$a\gamma$	$\dfrac{\Gamma(1-m_2)x_1^{m_1-1}x_2^{m_2-1}(-1-x_1+x_2)^{m_3-1}}{\Gamma(m_1)\Gamma(m_3)\Gamma(1-m_1-m_2-m_3)}$	$m_1, m_3 > 0$; $m_1+m_2+m_3 < 1$; $x_2 - 1 > x_1 > 0$	VI	VI
II$a\delta$	$\dfrac{\Gamma(1-m_1)x_1^{m_1-1}x_2^{m_2-1}(-1+x_1-x_2)^{m_3-1}}{\Gamma(m_2)\Gamma(m_3)\Gamma(1-m_1-m_2-m_3)}$	$m_1, m_3 > 0$; $m_1+m_2+m_3 < 1$; $x_1 - 1 > x_2 > 0$	VI	VI
IIb	$\dfrac{x_1^{m_1-1}x_2^{m_2-1}e^{-\frac{x_1+1}{x_2}}}{\Gamma(m_1)\Gamma(m_3)\Gamma(1-m_1-m_2-m_3)}$	$m_1 > 0$; $m_1 + m_2 < 0$; $x_1, x_2 > 0$	VI	V
III$a\alpha$	$\dfrac{m\sqrt{1-\rho^2}}{\pi k^m}\left(k+x_1^2+2\rho x_1x_2+x_2^2\right)^{m-1}$	$m < 0$; $\lvert\rho\rvert < 1$; $k > 0$	VII	VII
III$a\beta$	$\dfrac{m\sqrt{1-\rho^2}}{\pi k^m}\left(k-x_1^2+2\rho x_1x_2-x_2^2\right)^{m-1}$	$m < 0$; $\lvert\rho\rvert < 1$; $k > 0$; $x_1^2 - 2\rho x_1x_2 + x_2^2 < k$	II	II
IVa	$\dfrac{x_1^{m_1-1}(x_2-x_1)^{m_2-1}e^{-x_2}}{\Gamma(m_1)\Gamma(m_2)}$	$m_1, m_2 > 0$; $0 < x_1 < x_2$	III	III
VI	$\dfrac{1}{2\pi\sqrt{1-\rho^2}}\,e^{-\frac{x_1^2-2\rho x_1x_2+x_2^2}{2(1-\rho^2)}}$	$\lvert\rho\rvert < 1$	Normal	Normal

Table 3: Bivariate Pearson Surfaces (Elderton and Johnson, 1969)

For the bivariate $\mathrm{PIV}_2(\rho, \gamma_1, \gamma_2, \nu)$ distribution defined later in §2.3.2, (27) is the solution of the PDE (17)-(18)

$$\frac{\partial \ln f(x,y)}{\partial x} = \frac{E(x,y)}{F(x,y)} = \tag{17}$$

$$-\frac{(\nu+2)\left(x^3-\rho x^2 y+\nu(x-\rho y)\right)-\gamma_1\sqrt{\nu}\left(x^2-2\rho x y+y^2\right)+\gamma_1 \nu^{3/2}\left(\rho^2-1\right)}{\left(x^2-2\rho x y+y^2-\nu\rho^2+\nu\right)\left(\nu+x^2\right)}$$

$$\frac{\partial \ln f(x,y)}{\partial y} = \frac{G(x,y)}{H(x,y)} = \tag{18}$$

$$-\frac{(\nu+2)\left(y^3-\rho x y^2+\nu(y-\rho x)\right)-\gamma_2\sqrt{\nu}\left(x^2-2\rho x y+y^2\right)+\gamma_2 \nu^{3/2}\left(\rho^2-1\right)}{\left(x^2-2\rho x y+y^2-\nu\rho^2+\nu\right)\left(\nu+y^2\right)}$$

where the mixed partial condition

$$\frac{\partial^2 \ln f(x,y)}{\partial x\,\partial y} = -\frac{(\nu+2)\left(\rho\left(x^2+y^2-\nu\right)-2xy+\nu\rho^3\right)}{\left(x^2-2\rho x y+y^2-\nu\rho^2+\nu\right)^2} = \frac{\partial^2 \ln f(x,y)}{\partial y\,\partial x}$$

is a function of the association parameter ρ, and conveniently not dependent on the marginal skewness parameters γ_1 and γ_2. Note that (17)-(18) involve cubic polynomials $E(x,y)$ and $G(x,y)$ in the numerator and quartic polynomials $F(x,y)$ and $G(x,y)$ in the denominator whereas Van Uven's PDE only had linear $A(x,y)$ and $C(x,y)$, with quadratic $B(x,y)$ and $D(x,y)$ polynomials respectively. This explains why Van Uven was unable to discover a Pearson surface with marginal Type IV's, as a cubic polynomial is necessary to express the third moment (skewness) while a quartic polynomial is necessary to express the fourth moment (kurtosis).

2.2 Univariate Pearson IV Distribution

In this section, the special function $\mathcal{C}_1(\mu, \sigma, \gamma, \nu)$ is defined together with the univariate standard and location-scale Pearson IV distributions.

Lemma 2.3 states and proves a simple intuitive fact, ie. the volume under the special function integrand does not depend on the location parameter μ, and is inversely proportional to the scaling parameter σ.

Analytical expressions for the moments of this distribution are given in Theorem 2.5, and are identical when compared with results obtained using numerical integration (Figures 10-13).

2.2.1 $\mathcal{C}_1(\mu, \sigma, \gamma, \nu)$ special function

Definition 2.2 (univariate pearson iv special function) *Let $\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+, \gamma \in \mathbb{R}$, and $\nu \geq 1$. Then the univariate pearson iv special function (Figure 8) is defined as*

$$\mathcal{C}_1(\gamma, \nu) = \frac{1}{\int_{-\infty}^{\infty} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} e^{\gamma \tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right)} \, dx} \tag{19}$$

and the univariate location-scale pearson iv special function is defined as

$$\mathcal{C}_1(\mu, \sigma, \gamma, \nu) = \frac{1}{\int_{-\infty}^{\infty} \left(1 + \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} e^{\gamma \tan^{-1}\left(\frac{x-\mu}{\sigma\sqrt{\nu}}\right)} \, dx} \tag{20}$$

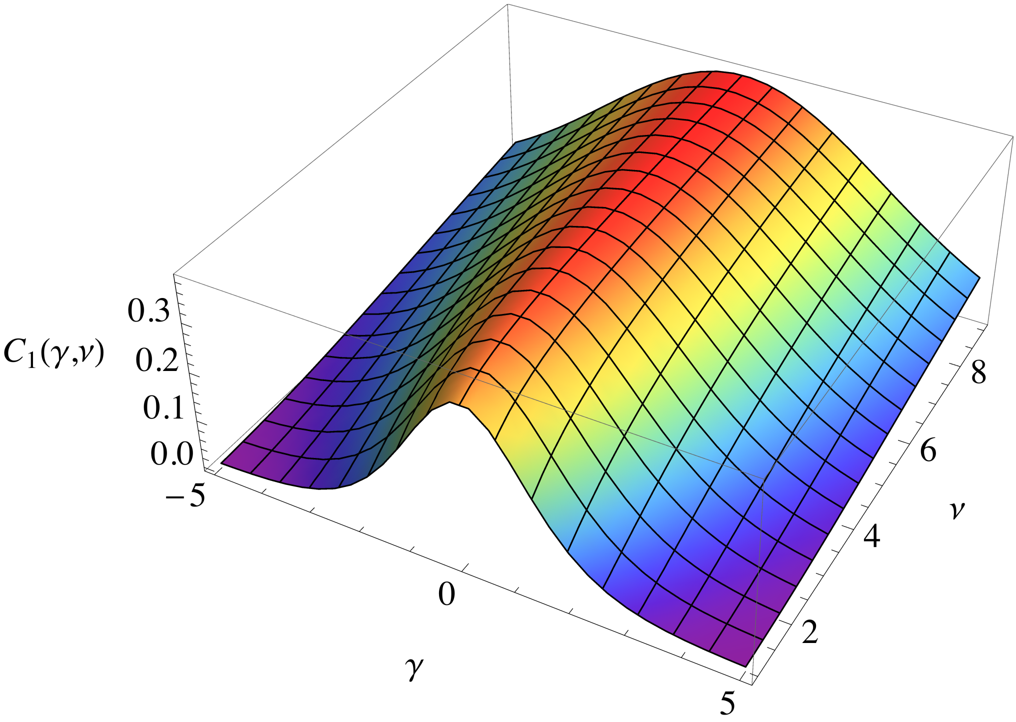


Figure 8: $\mathcal{C}_1(\gamma,\nu)$

Lemma 2.3 (univariate location-scale pearson iv special function) *The univariate location-scale pearson iv special function $\mathcal{C}_1(\mu,\sigma,\gamma,\nu)$ can be expressed in terms of the standard pearson iv special function $\mathcal{C}_1(\gamma,\nu)$, ie.*

$$\mathcal{C}_1(\mu,\sigma,\gamma,\nu) = \frac{\mathcal{C}_1(\gamma,\nu)}{\sigma} \tag{21}$$

Proof.

$$\mathcal{C}_1(\mu,\sigma,\gamma,\nu) = \frac{1}{\int_{-\infty}^{\infty}\left(1+\frac{\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} e^{\,\gamma\tan^{-1}\left(\frac{x-\mu}{\sigma\sqrt{\nu}}\right)}dx}$$

Substitute

$$z = \frac{x - \mu}{\sigma} \quad \Rightarrow \quad x = \sigma z + \mu \quad \Rightarrow \quad dx = \sigma dz$$

Next as $x \to -\infty, z \to -\infty$, and as $x \to \infty, z \to \infty$. Therefore

$$\mathcal{C}_1(\mu, \sigma, \gamma, \nu) = \frac{1}{\int_{-\infty}^{\infty} \left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}} e^{\gamma \tan^{-1}\left(\frac{z}{\sqrt{\nu}}\right)} (\sigma dz)}$$

$$= \frac{\mathcal{C}_1(\gamma, \nu)}{\sigma}$$

$\square$

2.2.2 PIV$_1(\gamma, \nu)$ and PIV$_1(\mu, \sigma, \gamma, \nu)$ distributions

Definition 2.4 (univariate standard and location-scale pearson iv distributions) *Let $\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+, \gamma \in \mathbb{R}$, and $\nu \geq 1$. Then the univariate standard Pearson IV distribution has pdf*

$$f(\, x \mid \gamma, \nu \,) \;=\; \mathcal{C}(\gamma, \nu) \, \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} e^{\gamma \tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right)}$$

and the univariate location-scale Pearson IV distribution has pdf

$$f(\, x \mid \mu, \sigma, \gamma, \nu \,) \;=\; \mathcal{C}(\mu, \sigma, \gamma, \nu) \, \left(1 + \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} e^{\gamma \tan^{-1}\left(\frac{x-\mu}{\sigma\sqrt{\nu}}\right)}$$

Theorem 2.5 (PIV$_1(\mu,\sigma,\gamma,\nu)$ moments) *The first four central moments (ie. mean, variance, skewness, kurtosis) of PIV$_1(\mu,\sigma,\gamma,\nu)$ can be expressed analytically:*

$$\mu_1 = \begin{cases} \mu + \dfrac{\sigma\gamma\sqrt{\nu}}{\nu-1}, & \nu > 1 \\ \infty, & \nu \leq 1 \end{cases}$$

$$\mu_2 = \begin{cases} \left(\dfrac{\sigma^2\nu}{\nu-2}\right)\left(1+\left(\dfrac{\gamma}{\nu-1}\right)^2\right), & \nu > 2 \\ \infty, & \nu \leq 2 \end{cases}$$

$$\mu_3 = \begin{cases} \dfrac{4\gamma}{\nu-3}\sqrt{\dfrac{\nu-2}{\gamma^2+(\nu-1)^2}}, & \nu > 3 \\ \infty, & \nu \leq 3 \end{cases}$$

$$\mu_4 = \begin{cases} \left(\dfrac{3(\nu-2)}{\nu-4}\right)\left(\dfrac{\gamma^2(\nu+5)+(\nu-3)(\nu-1)^2}{(\nu-3)\left(\gamma^2+(\nu-1)^2\right)}\right), & \nu > 4 \\ \infty, & \nu \leq 4 \end{cases}$$

Proof. Applying Lemma 2.1 with

$$a = -\mu - \frac{\sigma\gamma\sqrt{\nu}}{\nu+1}, \qquad b_0 = \frac{\sigma^2\nu+\mu^2}{\nu+1}, \qquad b_1 = -\frac{2\mu}{\nu+1}, \qquad b_2 = \frac{1}{\nu+1}$$

gives the raw moment recurrence relation

$$\mu'_n = \frac{\mu'_{n-1}\left(\mu(\nu-2n+1)+\gamma\sqrt{\nu}\sigma\right)+(n-1)\mu'_{n-2}\left(\mu^2+\nu\sigma^2\right)}{\nu-n}, \quad \mu'_{-1}=0,\ \mu'_0=1.$$

Mean

$$\mu_1 = \frac{\gamma\sqrt{\nu}\sigma + \mu(\nu-1)}{\nu-1} = \mu + \frac{\sigma\gamma\sqrt{\nu}}{\nu-1}$$

Variance

$$\mu_2 = \mu_2' - {\mu_1'}^2 = \frac{\frac{\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-3)\right)\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-1)\right)}{\nu-1} + \mu^2 + \nu\sigma^2}{\nu-2} - \frac{\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-1)\right)^2}{(\nu-1)^2} = \left(\frac{\sigma^2\nu}{\nu-2}\right)\left(1 + \left(\frac{\gamma}{\nu-1}\right)^2\right)$$

Skewness

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3$$

$$= \frac{\frac{2\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-1)\right)^3}{(\nu-1)^3} + \frac{\frac{2\left(\mu^2+\nu\sigma^2\right)\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-1)\right)}{\nu-1} + \frac{\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-5)\right)\left(\frac{\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-3)\right)\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-1)\right)}{\nu-1} + \mu^2 + \nu\sigma^2\right)}{\nu-2}}{\nu-3} - \frac{3\left(\frac{\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-3)\right)\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-1)\right)}{\nu-1} + \mu^2 + \nu\sigma^2\right)\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-1)\right)}{(\nu-2)(\nu-1)}}{\left(\frac{\frac{\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-3)\right)\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-1)\right)}{\nu-1} + \mu^2 + \nu\sigma^2}{\nu-2} - \frac{\left(\gamma\sqrt{\nu}\sigma + \mu(\nu-1)\right)^2}{(\nu-1)^2}\right)^{3/2}}$$

$$= \frac{4\gamma}{\nu-3}\sqrt{\frac{\nu-2}{\gamma^2 + (\nu-1)^2}}$$

Kurtosis

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$= \frac{(\nu-2)^2(\nu-1)^4}{\nu^2\sigma^4\left(\gamma^2+(\nu-1)^2\right)^2}\left(\frac{6\left(\frac{\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-3)\right)\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-1)\right)}{\nu-1}+\mu^2+\nu\sigma^2\right)\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-1)\right)^2}{(\nu-2)(\nu-1)^2}\right.$$

$$-\frac{4\left(\frac{2\left(\mu^2+\nu\sigma^2\right)\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-1)\right)}{\nu-1}+\frac{\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-5)\right)\left(\frac{\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-3)\right)\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-1)\right)}{\nu-1}+\mu^2+\nu\sigma^2\right)}{\nu-2}\right)\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-1)\right)}{(\nu-3)(\nu-1)}$$

$$+\frac{\frac{3\left(\mu^2+\nu\sigma^2\right)\left(\frac{\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-3)\right)\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-1)\right)}{\nu-1}+\mu^2+\nu\sigma^2\right)}{\nu-2}+\frac{\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-7)\right)\left(\frac{2\left(\mu^2+\nu\sigma^2\right)\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-1)\right)}{\nu-1}+\frac{\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-5)\right)\left(\frac{\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-3)\right)\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-1)\right)}{\nu-1}+\mu^2+\nu\sigma^2\right)}{\nu-2}\right)}{\nu-3}}{\nu-4}$$

$$\left.-\frac{3\left(\gamma\sqrt{\nu}\sigma+\mu(\nu-1)\right)^4}{(\nu-1)^4}\right)$$

$$= \left(\frac{3(\nu-2)}{\nu-4}\right)\left(\frac{\gamma^2(\nu+5)+(\nu-3)(\nu-1)^2}{(\nu-3)\left(\gamma^2+(\nu-1)^2\right)}\right).$$

$\square$

2.3 Bivariate Pearson IV Distribution

In this section, the special function $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$ is defined together with the bivariate standard and location-scale Pearson IV distributions. All definitions and results that follow are original.

2.3.1 $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$ special function

Definition 2.6 (bivariate pearson iv special function) *Let $\mu_1, \mu_2 \in \mathbb{R}, \sigma_1, \sigma_2 \in \mathbb{R}^+$, $\gamma_1, \gamma_2 \in \mathbb{R}$, and $\nu \geq 1$. Then the bivariate pearson iv special function is defined as*

$$\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu) = \frac{1}{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{\gamma_1 \tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y}{\sqrt{\nu}}\right)}}{\left(1 + \dfrac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}} \, dxdy} \tag{22}$$

and the bivariate location-scale pearson iv special function is defined as

$$\mathcal{C}_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2, \nu) =$$

$$\frac{1}{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{\gamma_1 \tan^{-1}\left(\frac{x-\mu_1}{\sigma_1\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y-\mu_2}{\sigma_2\sqrt{\nu}}\right)}}{\left(1 + \dfrac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}} \, dxdy} \tag{23}$$

Lemma 2.7 (bivariate location-scale pearson iv special function) *The bivariate location-scale pearson iv special function $\mathcal{C}_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2, \nu)$ can be expressed in terms of the bivariate standard pearson iv special function $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$, ie.*

$$\mathcal{C}_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2, \nu) = \frac{\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)}{\sigma_1 \sigma_2} \tag{24}$$

Proof.

$$\mathcal{C}_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2, \nu) =$$

$$\frac{1}{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{\gamma_1 \tan^{-1}\left(\frac{x-\mu_1}{\sigma_1\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y-\mu_2}{\sigma_2\sqrt{\nu}}\right)}}{\left(1 + \frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}}\, dxdy}$$

Substitute

$$w = \frac{x-\mu_1}{\sigma_1} \quad \Rightarrow \quad x = \sigma_1 w + \mu_1 \quad \Rightarrow \quad dx = \sigma_1 dw$$

and

$$z = \frac{y-\mu_2}{\sigma_2} \quad \Rightarrow \quad y = \sigma_2 z + \mu_2 \quad \Rightarrow \quad dy = \sigma_2 dz$$

Therefore

$$\mathcal{C}_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2, \nu)$$

$$= \frac{1}{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{\gamma_1 \tan^{-1}\left(\frac{w}{\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{z}{\sqrt{\nu}}\right)}}{\left(1 + \frac{w^2 - 2\rho wz + z^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}} (\sigma_1 dw)(\sigma_2 dz)}$$

$$= \frac{\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)}{\sigma_1 \sigma_2}$$

□

Lemma 2.8 (bivariate pearson iv special function symmetry) *The bivariate pearson iv special function $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$ has symmetry properties, ie.*

$$\mathcal{C}_2(-\rho, \gamma_1, \gamma_2, \nu) = \mathcal{C}_2(\rho, -\gamma_1, \gamma_2, \nu) = \mathcal{C}_2(\rho, \gamma_1, -\gamma_2, \nu) \tag{25}$$

that can be used to simplify the storage of cached values to speedup numerical computations.

Proof. Substituting $z = -x$,

$$
\begin{aligned}
\mathcal{C}_2(-\rho, \gamma_1, \gamma_2, \nu) &= \frac{1}{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{\gamma_1 \tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y}{\sqrt{\nu}}\right)}}{\left(1 + \frac{x^2 - 2(-\rho)xy + y^2}{\nu(1-(-\rho)^2)}\right)^{\frac{\nu+2}{2}}}\, dxdy} \\
&= \frac{1}{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{\gamma_1 \tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y}{\sqrt{\nu}}\right)}}{\left(1 + \frac{x^2 - 2\rho(-x)y + y^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}}\, dxdy} \\
&= \frac{1}{\displaystyle\int_{\infty}^{-\infty}\int_{-\infty}^{\infty} \frac{e^{\gamma_1 \tan^{-1}\left(\frac{-z}{\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y}{\sqrt{\nu}}\right)}}{\left(1 + \frac{(-z)^2 - 2\rho zy + y^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}}\, (-dz)dy} \\
&= \frac{1}{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{(-\gamma_1) \tan^{-1}\left(\frac{z}{\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y}{\sqrt{\nu}}\right)}}{\left(1 + \frac{z^2 - 2\rho zy + y^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}}\, dzdy} \\
&= \mathcal{C}_2(\rho, -\gamma_1, \gamma_2, \nu).
\end{aligned}
$$

and similarly, by substituting $z = -y$ instead,

$$
\mathcal{C}_2(-\rho, \gamma_1, \gamma_2, \nu) = \mathcal{C}_2(\rho, \gamma_1, -\gamma_2, \nu).
$$

□

2.3.2 PIV$_2(\gamma_1, \gamma_2, \nu)$, PIV$_2(\rho, \gamma_1, \gamma_2, \nu)$, and PIV$_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2, \nu)$ distributions

Definition 2.9 (bivariate Pearson IV distributions) *Let $\mu_1, \mu_2 \in \mathbb{R}, \sigma_1, \sigma_2 \in \mathbb{R}^+$, $\gamma_1, \gamma_2 \in \mathbb{R}$, and $\nu \geq 1$. The bivariate standard independent Pearson IV distribution has pdf*

$$f(x, y \mid \gamma_1, \gamma_2, \nu) = \mathcal{C}(\gamma_1, \gamma_2, \nu) \frac{e^{\gamma_1 \tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y}{\sqrt{\nu}}\right)}}{\left(1 + \frac{x^2 + y^2}{\nu}\right)^{\frac{\nu+2}{2}}} \tag{26}$$

the bivariate standard dependent Pearson IV distribution (Figure 9) has pdf

$$f(x, y \mid \rho, \gamma_1, \gamma_2, \nu) =$$

$$\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu) \frac{e^{\gamma_1 \tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y}{\sqrt{\nu}}\right)}}{\left(1 + \frac{x^2 - 2\rho x y + y^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}} \tag{27}$$

and the bivariate location-scale dependent Pearson IV distribution has pdf

$$f(x, y \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2, \nu) =$$

$$\frac{\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)}{\sigma_1 \sigma_2} \frac{e^{\gamma_1 \tan^{-1}\left(\frac{x-\mu_1}{\sigma_1\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{y-\mu_2}{\sigma_2\sqrt{\nu}}\right)}}{\left(1 + \frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}} \tag{28}$$

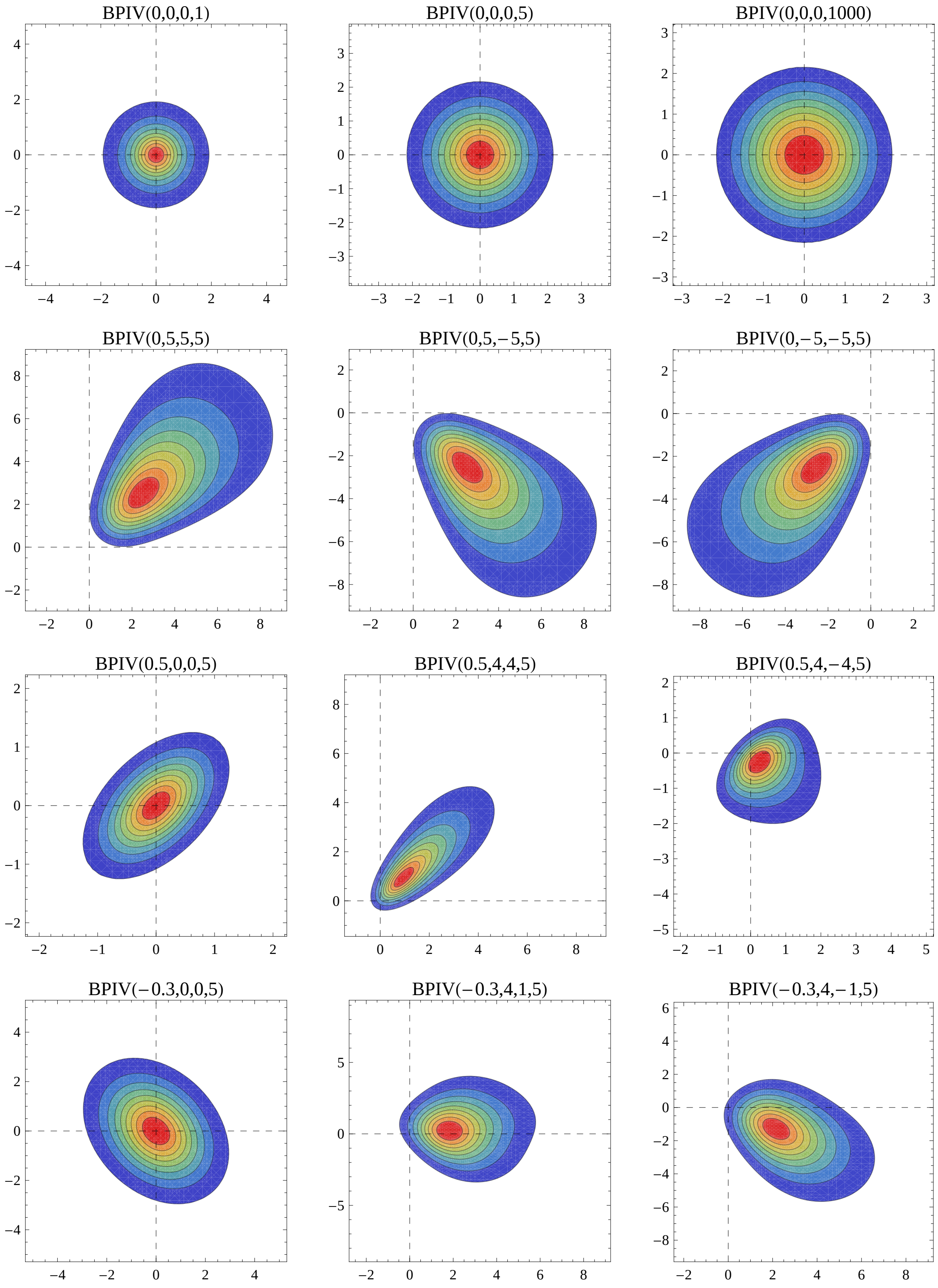


Figure 9: Bivariate Pearson IV Distribution - BPIV($\rho, \gamma_1, \gamma_2, \nu$) plots

2.4 Multivariate Pearson IV Distribution

In this section, the special function $\mathcal{C}_N(\boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu)$ is defined together with the multivariate Pearson IV distribution. Note that (30) is simply a multivariate t distribution (Kotz and Nadarajah, 2004) multiplied by $e^{\sum_{i=1}^{N} \gamma_i \tan^{-1}\left(\frac{x_i-\mu_i}{\sigma_i\sqrt{\nu}}\right)}$.

2.4.1 $\mathcal{C}_N(\boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu)$ special function

Definition 2.10 (bivariate pearson iv special function) *Let $\boldsymbol{\Delta}$ be a covariance matrix, $\vec{\boldsymbol{\gamma}} \in \mathbb{R}^N$, and $\nu \geq 1$. Then the multivariate Pearson IV special function is defined as*

$$\mathcal{C}_N(\boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu) = \frac{1}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(1 + \frac{\vec{\boldsymbol{x}}^T \boldsymbol{\Delta}^{-1} \vec{\boldsymbol{x}}}{\nu}\right)^{-\frac{\nu+N}{2}} e^{\sum_{i=1}^{N} \gamma_i \tan^{-1}\left(\frac{x_i}{\sigma_i\sqrt{\nu}}\right)} \, d\vec{\boldsymbol{x}}} \tag{29}$$

2.4.2 $\mathrm{PIV}_N(\vec{\boldsymbol{\mu}}, \boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu)$ distribution

Definition 2.11 (multivariate pearson iv distribution) *Let $\vec{\boldsymbol{\mu}} \in \mathbb{R}^N, \vec{\boldsymbol{\gamma}} \in \mathbb{R}^N, \nu \geq 1$, and $\boldsymbol{\Delta}$ be a covariance matrix. Then the multivariate location-scale Pearson IV distribution has pdf*

$$f(\vec{\boldsymbol{x}} \mid \vec{\boldsymbol{\mu}}, \boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu) = \mathcal{C}_N(\boldsymbol{\Delta}, \vec{\boldsymbol{\gamma}}, \nu) \frac{e^{\sum_{i=1}^{N} \gamma_i \tan^{-1}\left(\frac{x_i-\mu_i}{\sigma_i\sqrt{\nu}}\right)}}{\left(1 + \frac{(\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})^T \boldsymbol{\Delta}^{-1} (\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})}{\nu}\right)^{\frac{\nu+N}{2}}} \tag{30}$$

Dilbert's boss: *"You spelled that word wrong."*

Dilbert: *"That's a number!"*

3 Numerics

Numerical methods presented in this section, and in general, are essential tools of an applied mathematician. Analytical expressions sometimes exist but are computationally expensive or rely on obscure functions missing from standard numerical libraries, eg. the complex beta function for the univariate $\mathcal{C}_1(\gamma, \nu)$ special function Willink (2008) representation. In most cases, however, numerical approximations are the only available methods to obtain results.

Numerical integration, both univariate and multivariate, estimates the value of an integral either by sampling the integrand randomly (Monte Carlo methods) or using deterministic rules until a desired accuracy is reached.

Maximum likelihood estimation, an essential procedure when dealing with distributions, would be near impossible without numerical optimization algorithms that can reach an estimate by successive refinements towards the correct answer.

During numerical maximum likelihood, values of the special functions $\mathcal{C}_1(\gamma, \nu)$ in the univariate case, and $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$ in the bivariate case, are needed at each step of the algorithm. *Multivariate interpolation* on a grid of precomputed values can significantly improve performance.

3.1 Numerical Integration

3.1.1 Univariate

Table 4 gives values obtained using QUADPACK (Piessens et al., 1983) for the $\mathcal{C}_1(\gamma,\nu)$ special function to *IEEE float precision*, ie. relative error of 10^{-8}. First, the $\mathcal{C}_1(\gamma,\nu)$ integral over $\mathbb{R}$ needs to be transformed to $[0,1]$ using the following lemma.

Lemma 3.1 (univariate pearson iv special function numerical representation) *The univariate pearson iv special function $\mathcal{C}_1(\gamma,\nu)$ can be expressed as an integral over* $[0,1]$*, suitable for numerical integration algorithms, ie.*

$$\mathcal{C}_1(\gamma,\nu) = \frac{e^{\gamma\pi/2}}{\pi\sqrt{\nu}\int_0^1 \left(1+\cot^2(\pi\theta)\right)^{-\frac{\nu+1}{2}} e^{\gamma\pi\theta}\csc^2(\pi\theta)\,d\theta} \tag{31}$$

Proof. By Definition 2.2,

$$\mathcal{C}_1(\gamma,\nu) = \frac{1}{\int_{-\infty}^{\infty}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} e^{\gamma\tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right)}\,dx}$$

Subtitute $\theta = \dfrac{1}{2} + \dfrac{\tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right)}{\pi}$

$$\Rightarrow \quad \tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right) = \pi\left(\theta - \frac{1}{2}\right)$$

$$\Rightarrow \quad x = \sqrt{\nu}\tan\left[\pi\left(\theta - \frac{1}{2}\right)\right] = -\sqrt{\nu}\cot(\pi\theta)$$

$$\Rightarrow \quad dx = \pi\sqrt{\nu}\csc^2(\pi\theta)\,d\theta$$

Next as $x \to -\infty$, $\tan^{-1}\left(x/\sqrt{\nu}\right) \to -\frac{\pi}{2}$, so $\theta \to \frac{1}{2} + \frac{-\cancel{\pi}/2}{\cancel{\pi}} = 0$ and similarly as $x \to \infty$, $\theta \to \frac{1}{2} + \frac{\cancel{\pi}/2}{\cancel{\pi}} = 1$. Therefore

$$\mathcal{C}_1(\gamma,\nu) = \frac{1}{\int_{-\infty}^{\infty}\left(1 + \frac{\cancel{\nu}\cot^2(\pi\theta)}{\cancel{\nu}}\right)^{-\frac{\nu+1}{2}} e^{\gamma\pi\left(\theta-\frac{1}{2}\right)}\left(\pi\sqrt{\nu}\csc^2(\pi\theta)\,d\theta\right)}$$

$$= \frac{e^{\gamma\pi/2}}{\pi\sqrt{\nu}\int_0^1\left(1+\cot^2(\pi\theta)\right)^{-\frac{\nu+1}{2}} e^{\gamma\pi\theta}\csc^2(\pi\theta)\,d\theta}$$

□

Genz and Kass (1997) showed that a subregion-adaptive algorithm, together with a transformation based on location of the dominant peak of the integrand, is considerably more efficient than Monte Carlo importance sampling for the univariate Pearson IV distribution.

γ	$\nu = 1$	$\nu = 4$	$\nu = 6$	$\nu = 8$	$\nu = 10$	$\nu = 100$
0	.31830988	.37500000	.38273277	.38669902	.38910838	.39794618
.1	.31700464	.37454061	.38241681	.38645864	.38891452	.39792629
.2	.31313356	.37316635	.38147069	.38573849	.38833354	.39786660
.3	.30682714	.37088884	.37989957	.38454146	.38736731	.39776715
.4	.29829198	.36772723	.37771202	.38287237	.38601889	.39762796
.5	.28779596	.36370793	.37491993	.38073789	.38429255	.39744908
.6	.27565026	.35886411	.37153835	.37814655	.38219375	.39723055
.7	.26219012	.35323525	.36758540	.37510862	.37972908	.39697245
.8	.24775655	.34686650	.36308202	.37163607	.37690626	.39667485
.9	.23268021	.33980796	.35805186	.36774249	.37373408	.39633783
1	.21726860	.33211401	.35252094	.36344296	.37022233	.39596152
1.1	.20179693	.32384245	.34651750	.35875399	.36638179	.39554600
1.2	.18650261	.31505375	.34007169	.35369335	.36222413	.39509142
1.3	.17158289	.30581022	.33321531	.34828004	.35776189	.39459791
1.4	.15719514	.29617522	.32598152	.34253404	.35300835	.39406561
1.5	.14345907	.28621236	.31840456	.33647630	.34797755	.39349469
1.6	.13046015	.27598483	.31051945	.33012853	.34268412	.39288531
1.7	.11825387	.26555466	.30236173	.32351307	.33714329	.39223767
1.8	.10687021	.25498212	.29396715	.31665279	.33137075	.39155195
1.9	.09631808	.24432520	.28537143	.30957091	.32538261	.39082835
2	.08658953	.23363908	.27660999	.30229090	.31919530	.39006710
2.2	.06950908	.21238373	.25872860	.28723068	.30629015	.38843255
2.4	.05535928	.19158849	.24059141	.27165950	.29279047	.38665023
2.6	.04379325	.17156326	.22244834	.25576038	.27883185	.38472227
2.8	.03444267	.15255291	.20452577	.23970851	.26454801	.38265094
3	.02695204	.13473849	.18702298	.22366823	.25006868	.38043868
3.5	.01433562	.09610094	.14614309	.18457197	.21380954	.37430887
4	.00746979	.06613836	.11070358	.14837351	.17878892	.36735943
4.5	.00383147	.04410181	.08147031	.11633235	.14632146	.35964027
5	.00194101	.02859763	.05837467	.08907124	.11728861	.35120601
10	.00000150	.00013825	.00070553	.00216763	.00487391	.24157097

Table 4: $\mathcal{C}_1(\gamma, \nu)$ - table

The first four moments, ie. the mean (Figure 10)

$$\mu_1 = \text{Mean}\left(PIV_1(\mu,\sigma,\gamma,\nu)\right) = \int_{-\infty}^{\infty} x \; f(\, x \mid \mu,\sigma,\gamma,\nu \,) \; dx$$

variance (Figure 11)

$$\mu_2 = \text{Var}\left(PIV_1(\mu,\sigma,\gamma,\nu)\right) = \int_{-\infty}^{\infty} (x-\mu_1)^2 \; f(\, x \mid \mu,\sigma,\gamma,\nu \,) \; dx$$

skewness (Figure 12)

$$\mu_3 = \pm\sqrt{\beta_1} = \text{Skewness}\left(PIV_1(\mu,\sigma,\gamma,\nu)\right) = \frac{1}{\mu_2^3}\int_{-\infty}^{\infty} (x-\mu_1)^3 \; f(\, x \mid \mu,\sigma,\gamma,\nu \,) \; dx$$

and kurtosis (Figure 13)

$$\mu_4 = \beta_2 = \text{Kurtosis}\left(PIV_1(\mu,\sigma,\gamma,\nu)\right) = \frac{1}{\mu_2^4}\int_{-\infty}^{\infty} (x-\mu_1)^4 \; f(\, x \mid \mu,\sigma,\gamma,\nu \,) \; dx$$

can be computed numerically and compared with their analytical expressions from Theorem 2.5.

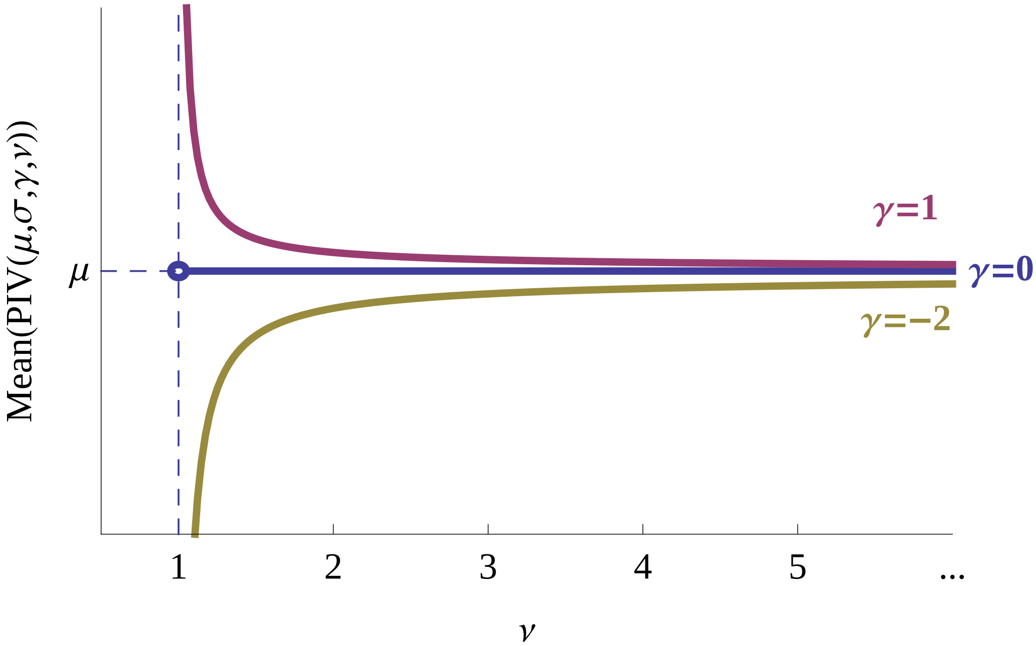


Figure 10: Mean $(PIV(\mu, \sigma, \gamma, \nu))$

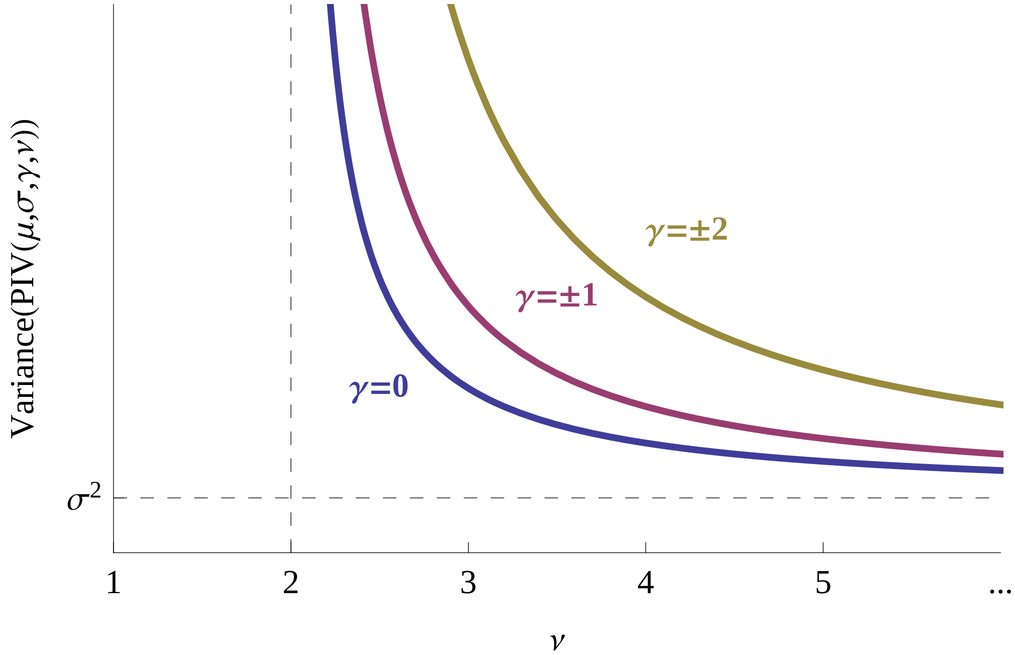


Figure 11: Var $(PIV(\mu, \sigma, \gamma, \nu))$

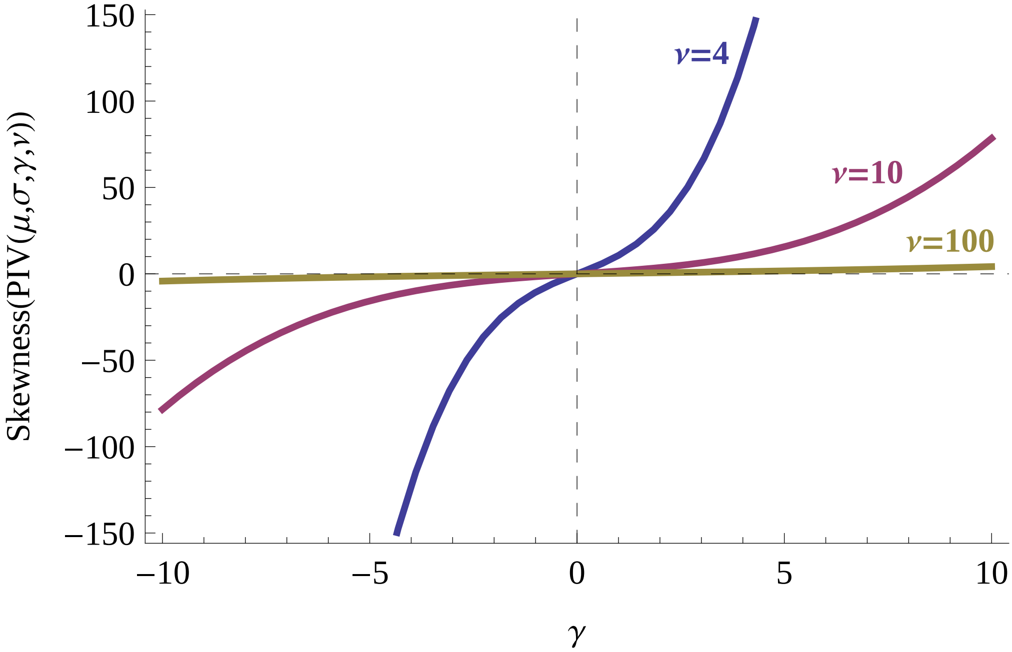


Figure 12: Skewness $(PIV(\mu, \sigma, \gamma, \nu))$

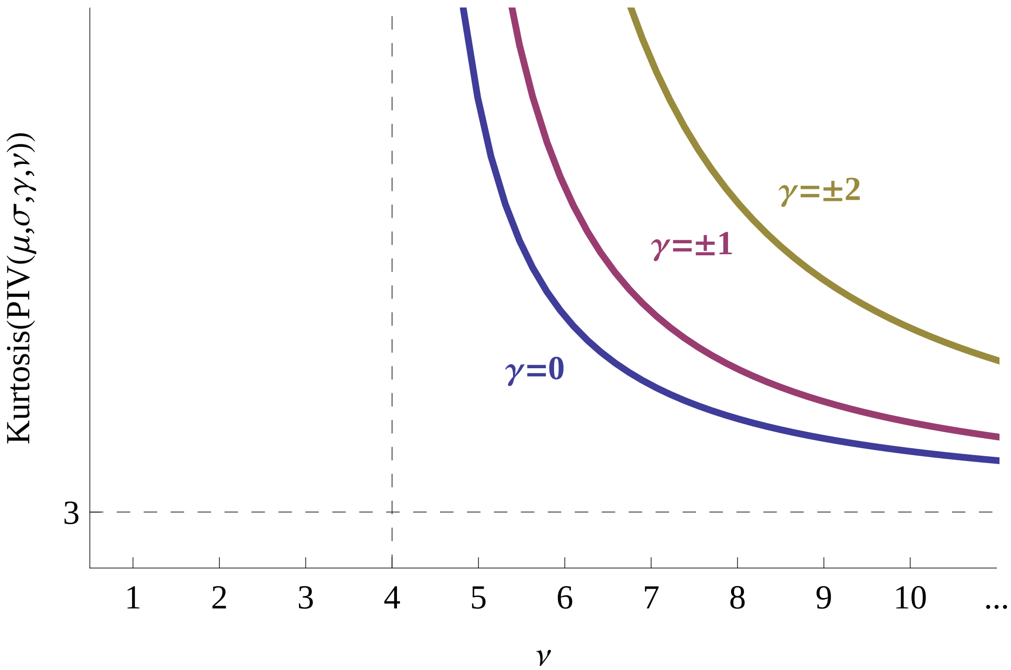


Figure 13: Kurtosis $(PIV(\mu, \sigma, 0, \nu))$

3.1.2 Multivariate

The CUHRE algorithm (Genz and Malik, 1980a,b, 1983; Berntsen et al., 1991) *"employs a cubature rule for subregion estimation in a globally adaptive subdivision scheme. It is hence a deterministic, not a Monte Carlo method. In each iteration, the subregion with the largest error is halved along the axis where the integrand has the largest fourth difference"* (Hahn, 2005, p.79). As demonstrated by Figure 15, *"CUHRE is quite powerful in moderate dimensions and is usually the only viable method to obtain high precision, ie. relative accuracies below* 10^{-3}*"* (Hahn, 2005, p.79). Numerical tests (Genz and Bretz, 2002; Hong and Hickernell, 2003) of this subregion adaptive method have shown it to perform well when computing multivariate normal and t probabilities in dimensions $N \leq 10$. However, Gassmann et al. (2002) detected numerical instability issues when using this algorithm for computing multivariate normal probabilities with ill-conditioned correlation matrices.

Originally published in FORTRAN (Genz and Cools, 2003), a C implementation is now available (Hahn, 2005) which also includes the Monte Carlo algorithms:

- *VEGAS* (Lepage et al., 1978; Lepage, 1980; Bratley and Fox, 1988; Niederreiter, 1992; Press et al., 1993);

- *SUAVE*, new and similar to recursive stratified sampling *MISER* (Press and Farrar, 1990);

- *DIVONNE* (Friedman and Wright, 1981) with a choice of methods for the integral estimates:
  - korobov (Korobov, 1963; Keng and Yuan, 1981);
  - sobol (Bratley and Fox, 1988);
  - mersenne twister (Matsumoto and Nishimura, 1998); or
  - same cubature rules of Genz and Malik (1983) used in CUHRE.

Similarly to the univariate case, the following lemma transforms the bivariate $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$ special function integrand from $\mathbb{R}^2$ to $[0, 1]^2$, as needed by the CUBA library. As seen in Figure 14, the transformed integrand in $[0, 1]^2$ is also well-behaved.

Lemma 3.2 (bivariate pearson iv special function numerical representation) *The bivariate pearson iv special function $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$ can be expressed as an integral over $[0, 1]^2$, suitable for numerical integration algorithms, ie.*

$$\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu) =$$

$$\frac{e^{\frac{\pi}{2}(\gamma_1+\gamma_2)}}{\pi^2\nu \displaystyle\int_0^1\int_0^1 \frac{e^{\pi(\gamma_1\theta+\gamma_2\phi)}\csc^2(\pi\theta)\csc^2(\pi\phi)}{\left(1+\dfrac{\cot^2(\pi\theta)-2\rho\cot(\pi\theta)\cot(\pi\phi)+\cot^2(\pi\phi)}{1-\rho^2}\right)^{\frac{\nu+2}{2}}}d\theta d\phi} \tag{32}$$

Proof. Substitute

$$\theta = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right), \qquad \phi = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{y}{\sqrt{\nu}}\right)$$

$$\Rightarrow \quad \tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right) = \pi\left(\theta - \frac{1}{2}\right) \;\Rightarrow\; x = -\sqrt{\nu}\cot(\pi\theta) \;\Rightarrow\; dx = \pi\sqrt{\nu}\csc^2(\pi\theta)d\theta$$

and similarly

$$\tan^{-1}\left(\frac{y}{\sqrt{\nu}}\right) = \pi\left(\phi - \frac{1}{2}\right) \;\Rightarrow\; y = -\sqrt{\nu}\cot(\pi\phi) \;\Rightarrow\; dy = \pi\sqrt{\nu}\csc^2(\pi\phi)d\phi$$

Next as $x \to -\infty$, $\tan^{-1}(x/\sqrt{\nu}) \to -\frac{\pi}{2}$, so $\theta \to \frac{1}{2} + \frac{-\cancel{\pi}/2}{\cancel{\pi}} = 0$ and as $x \to \infty$, $\theta \to \frac{1}{2} + \frac{\cancel{\pi}/2}{\cancel{\pi}} = 1$. Similarly as $y \to -\infty, \phi \to 0$, and as $y \to \infty, \phi \to 1$. Therefore

$\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$

$$= \frac{1}{\displaystyle\int_0^1\int_0^1 \frac{e^{\gamma_1\pi\left(\theta-\frac{1}{2}\right)+\gamma_2\pi\left(\phi-\frac{1}{2}\right)}\left(\pi\sqrt{\nu}\csc^2(\pi\theta)d\theta\right)\left(\pi\sqrt{\nu}\csc^2(\pi\phi)d\phi\right)}{\left[1+\frac{\cancel{\nu}\cot^2(\pi\theta)-2\rho\cancel{\sqrt{\nu}}\cot(\pi\theta)\cancel{\sqrt{\nu}}\cot(\pi\phi)+\cancel{\nu}\cot^2(\pi\phi)}{\cancel{\nu}(1-\rho^2)}\right]^{\frac{\nu+2}{2}}}}$$

$$= \frac{e^{\frac{\pi}{2}(\gamma_1+\gamma_2)}}{\displaystyle\pi^2\nu\int_0^1\int_0^1 \frac{e^{\pi(\gamma_1\theta+\gamma_2\phi)}\csc^2(\pi\theta)\csc^2(\pi\phi)}{\left(1+\frac{\cot^2(\pi\theta)-2\rho\cot(\pi\theta)\cot(\pi\phi)+\cot^2(\pi\phi)}{1-\rho^2}\right)^{\frac{\nu+2}{2}}}d\theta d\phi}$$

$\square$

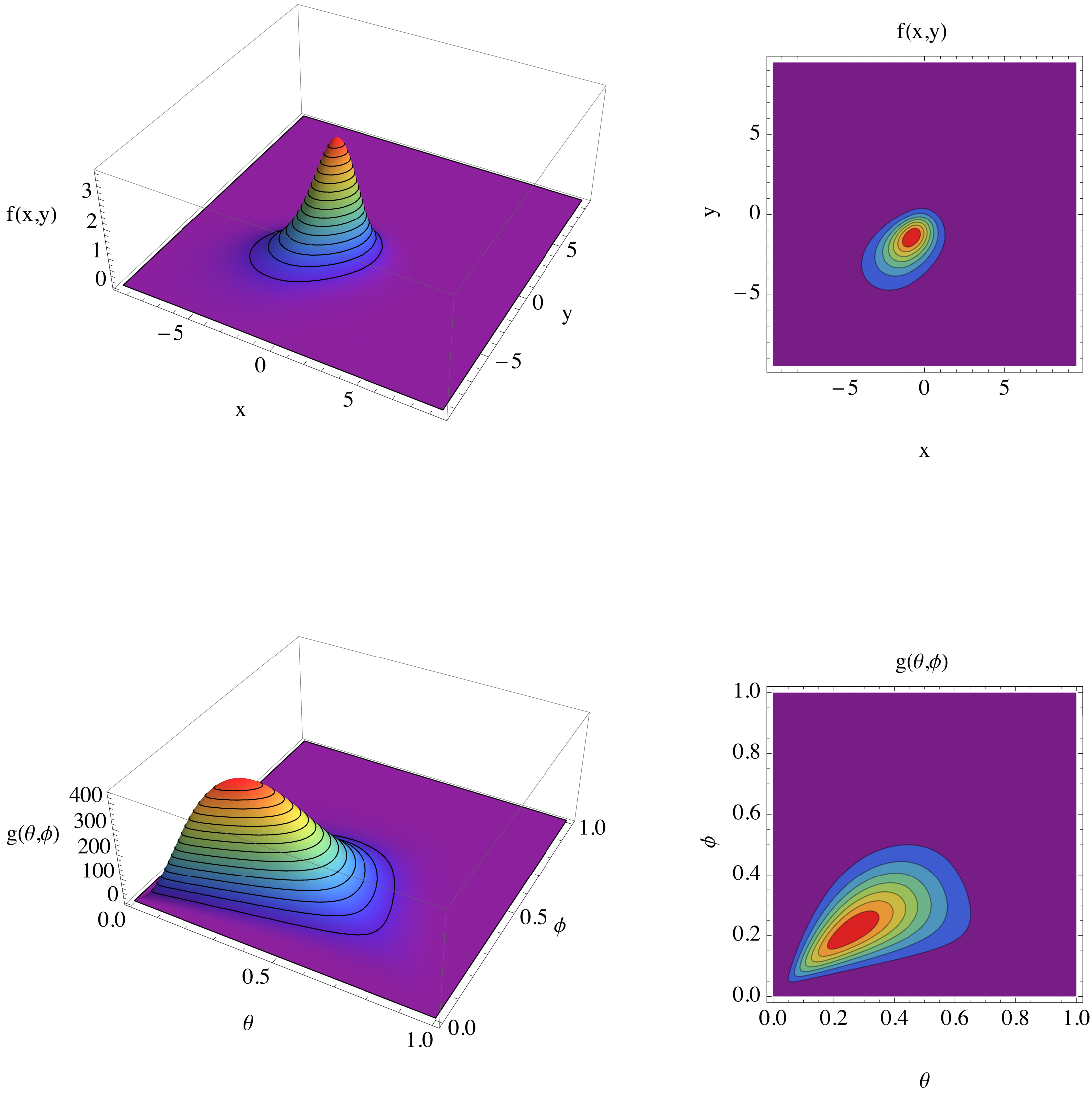


Figure 14: BPIV$(.3, -1, -4, 5)$ surface in $(x, y) \in \mathbb{R}^2$ and $(\theta, \phi) \in [0, 1]^2$ coordinates

Figure 15 compares the performance of CUHRE, VEGAS, SUAVE, and DIVONNE. CUHRE is showing much faster convergence than VEGAS, SUAVE, and DIVONNE, all of which do not achieve relative error much lower than 10^{-5}. Table 5 gives values obtained using CUHRE for the $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$ special function to *IEEE float precision*.

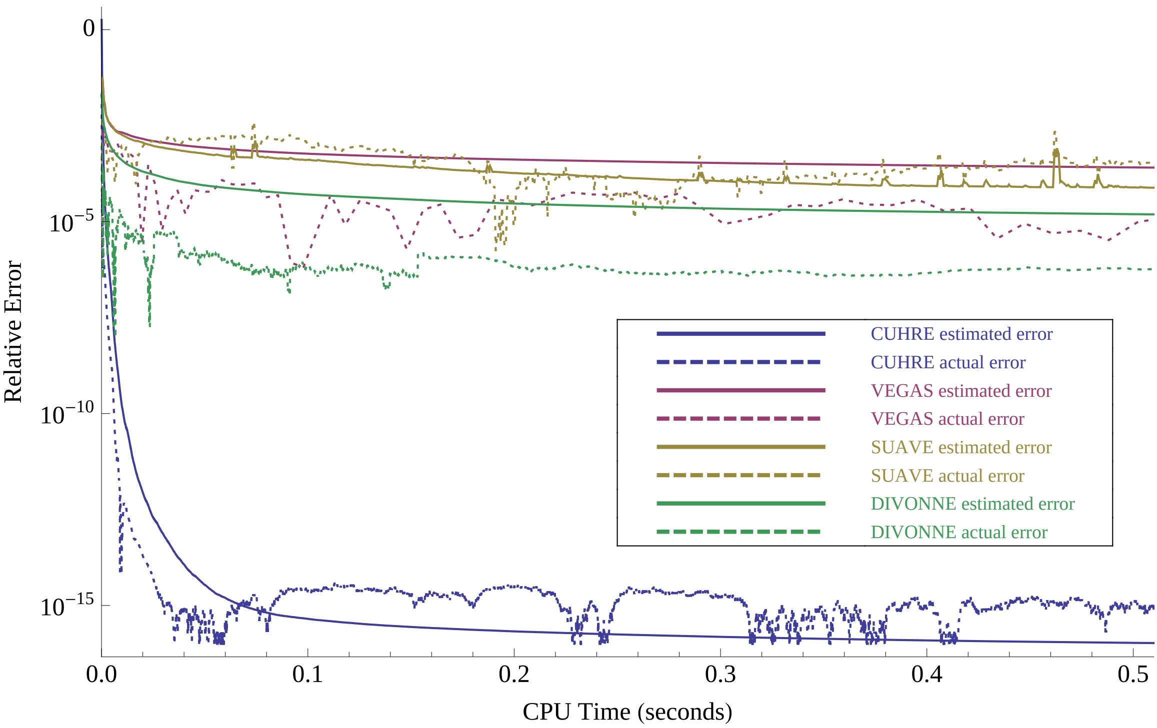


Figure 15: CUHRE/VEGAS/SUAVE/DIVONNE Convergence Diagram

ρ	γ_1	γ_2	$\nu = 1$	$\nu = 4$	$\nu = 6$	$\nu = 8$	$\nu = 10$	$\nu = 100$
0	0	0	.15915494	.15915494	.15915494	.15915494	.15915494	.15915494
0	0	.5	.14389798	.15436244	.15590606	.15670150	.15718515	.15895613
0	0	1	.10863430	.14095356	.14659170	.14958337	.15143008	.15836119
0	0	2	.04329476	.09915950	.11502502	.12441482	.13055876	.15600377
0	.5	0	.14389798	.15436244	.15590606	.15670150	.15718515	.15895613
0	.5	.5	.12954615	.14966102	.15270462	.15427729	.15523517	.15875756
0	.5	1	.09677312	.13651796	.14352872	.14724479	.14953832	.15816335
0	.5	2	.03763244	.09566818	.11246090	.12238942	.12888267	.15580882
0	1	0	.10863430	.14095356	.14659170	.14958337	.15143008	.15836119
0	1	.5	.09677312	.13651796	.14352872	.14724479	.14953832	.15816335
0	1	1	.07048472	.12414890	.13475670	.14046304	.14401256	.15757132
0	1	2	.02592251	.08602658	.10514115	.11652404	.12399037	.15522543
0	2	0	.04329476	.09915950	.11502502	.12441482	.13055876	.15600377
0	2	.5	.03763244	.09566818	.11246090	.12238942	.12888267	.15580882
0	2	1	.02592251	.08602658	.10514115	.11652404	.12399037	.15522543
0	2	2	.00853871	.05724904	.08070887	.09592405	.10631010	.15291377
.25	0	0	.16437451	.16437451	.16437451	.16437451	.16437451	.16437451
.25	0	.5	.14861719	.15942484	.16101909	.16184061	.16234012	.16416918
.25	0	1	.11219702	.14557621	.15139926	.15448904	.15639631	.16355473
.25	0	2	.04471464	.10241149	.11879733	.12849507	.13484051	.16112000
.25	.5	0	.14861719	.15942484	.16101909	.16184061	.16234012	.16416918
.25	.5	.5	.12859423	.15231018	.15612352	.15811679	.15933764	.16386167
.25	.5	1	.09315037	.13695497	.14527866	.14975968	.15254672	.16314637
.25	.5	2	.03499508	.09344295	.11163526	.12261661	.12987687	.16051691
.25	1	0	.11219702	.14557621	.15139926	.15448904	.15639631	.16355473
.25	1	.5	.09315037	.13695497	.14527866	.14975968	.15254672	.16314637
.25	1	1	.06447395	.12110168	.13371808	.14070456	.14511459	.16233267
.25	1	2	.02254518	.07972951	.10041276	.11328802	.12193587	.15951668
.25	2	0	.04471464	.10241149	.11879733	.12849507	.13484051	.16112000
.25	2	.5	.03499508	.09344295	.11163526	.12261661	.12987687	.16051691
.25	2	1	.02254518	.07972951	.10041276	.11328802	.12193587	.15951668
.25	2	2	.00704446	.04846630	.07166260	.08797686	.09965055	.15635760
.5	0	0	.18377629	.18377629	.18377629	.18377629	.18377629	.18377629
.5	0	.5	.16615907	.17824239	.18002481	.18094331	.18150177	.18354673
.5	0	1	.12544008	.16275915	.16926952	.17272400	.17485639	.18285975
.5	0	2	.04999249	.11449953	.13281946	.14366186	.15075627	.18013764
.5	.5	0	.16615907	.17824239	.18002481	.18094331	.18150177	.18354673
.5	.5	.5	.13816513	.16779364	.17279028	.17542512	.17704584	.18308846
.5	.5	1	.09721533	.14875842	.15919641	.16489348	.16846160	.18217535
.5	.5	2	.03541633	.09893347	.12002841	.13301940	.14170218	.17901536
.5	1	0	.12544008	.16275915	.16926952	.17272400	.17485639	.18285975
.5	1	.5	.09721533	.14875842	.15919641	.16489348	.16846160	.18217535
.5	1	1	.06430947	.12802431	.14370067	.15261012	.15831069	.18104036
.5	1	2	.02157010	.08039969	.10403483	.11935758	.12988771	.17745584
.5	2	0	.04999249	.11449953	.13281946	.14366186	.15075627	.18013764
.5	2	.5	.03541633	.09893347	.12002841	.13301940	.14170218	.17901536
.5	2	1	.02157010	.08039969	.10403483	.11935758	.12988771	.17745584
.5	2	2	.00646227	.04517351	.06943882	.08770811	.10135180	.17307455
.75	0	0	.24061965	.24061965	.24061965	.24061965	.24061965	.24061965
.75	0	.5	.21755330	.23337407	.23570781	.23691040	.23764161	.24031908
.75	0	1	.16423962	.21310175	.22162582	.22614880	.22894076	.23941962
.75	0	2	.06545553	.14991508	.17390149	.18809753	.19738630	.23585553
.75	.5	0	.21755330	.23337407	.23570781	.23691040	.23764161	.24031908
.75	.5	.5	.17342830	.21643017	.22393671	.22791774	.23037343	.23956930
.75	.5	1	.11854508	.18919569	.20428456	.21261486	.21786313	.23822560
.75	.5	2	.04196958	.12278697	.15120726	.16903928	.18108370	.23380108
.75	1	0	.16423962	.21310175	.22162582	.22614880	.22894076	.23941962
.75	1	.5	.11854508	.18919569	.20428456	.21261486	.21786313	.23822560
.75	1	1	.07514072	.15857594	.18089134	.19386559	.20226878	.23644575
.75	1	2	.02429657	.09531061	.12646502	.14741365	.16212322	.23118595
.75	2	0	.06545553	.14991508	.17390149	.18809753	.19738630	.23585553
.75	2	.5	.04196958	.12278697	.15120726	.16903928	.18108370	.23380108
.75	2	1	.02429657	.09531061	.12646502	.14741365	.16212322	.23118595
.75	2	2	.00701617	.04992567	.07933919	.10279552	.12099604	.22435487

Table 5: $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$ - table

As discussed in Genz and Bretz (2009, p.69), multivariate integration *"involve computationally intensive but repetitive evaluation of integrands"* which can be computed in parallel at each step of the algorithms.

Hajivassiliou (1993) compares the impact of vectorization on computational performance of 13 multivariate integration algorithms on the CRAY-Y/MP4 vector supercomputer. In some cases, a speedup of more than $10X$ was observed, while in others performance gains were negligible.

Using the highly-portable *Message Passing Interface (MPI)* parallel programming environment, de Doncker et al. (1999, 2001) implement *asynchronous sampling* for multivariate normal and t probabilities, which are available as part of the *ParInt* parallel integration library (Dedoncker et al., 1996; de Doncker et al., 2002).

This can easily be extended to MPIV computations, by replacing the sampling function (Figure 16) from the Hahn (2005) CUHRE implementation with a parallelized version (Figure 17) written for the MAC OS X Grand Central Dispatch library. Unfortunately, only minor improvements in performance were achieved due to the high overhead of creating/destroying many threads (Table 6).

time (%)	library	function
74.3%	libSystem.B.dylib	__workq_kernreturn
14.6%	mpiv_multicore	Integrate
3.1%	libSystem.B.dylib	start_wqthread
1.3%	libSystem.B.dylib	semaphore_wait_trap
1.0%	libSystem.B.dylib	pow$fenv_access_off
0.9%	libSystem.B.dylib	tan$fenv_access_off
0.6%	mpiv_multicore	bpiv_sf_integrand

Table 6: Parallelized CUHRE time profiling

```
static inline void DoSample(This *t, count n, creal *x, real *f) {

    t->neval += n;

    while( n-- ) {

        if( t->integrand(&t->ndim, x, &t->ncomp, f, t->userdata) == ABORT )
            longjmp(t->abort, 1);

        x += t->ndim;
        f += t->ncomp;

    }

}
```

Figure 16: CUHRE DoSample() function

```
#include <dispatch/dispatch.h>

static inline void DoSample(This *t, count n, creal *x, real *f) {

    t->neval += n;
    dispatch_queue_t queue = dispatch_get_global_queue(DISPATCH_QUEUE_PRIORITY_DEFAULT, 0);

    dispatch_apply( n, queue, ^(size_t index) {
        if( t->integrand(&t->ndim, &x[index*t->ndim], &t->ncomp, &f[index*t->ncomp], t->userdata) == ABORT )
            longjmp(t->abort, 1);
    });

}
```

Figure 17: CUHRE DoSample() parallelized function

3.2 Maximum Likelihood Estimation

"[The fitting of Pearson distributions by] the method of moments [...] does not, in general, lead to efficient[‡] *estimates of the population parameters"* (Stuart and Ord, 1994, p.225-226). A more effective approach is to estimate the distribution parameters by maximum likelihood[§]. The maximum likelihood estimates (MLE) can be obtained by numerical maximization of the likelihood functions described in this section. Many of these algorithms, however, require either a gradient or partial derivatives in multiple directions to be computed. In the current context where analytical derivatives of the likelihood functions are not available, they would need to be approximated using centered-differences resulting in a considerable computational inefficiency.

The Nelder and Mead (1965) simplex algorithm does not require the gradient of the function to be computed. This algorithm *"maintains $n+1$ trial parameter vectors as the vertices of a n-dimensional simplex. On each iteration it tries to improve the worst vertex of the simplex by geometrical transformations. The iterations are continued until the overall size of the simplex has decreased sufficiently. [...] Using these transformations the simplex moves through the space towards the minimum, where it contracts itself"* (Galassi et al., 2010, p.393-399).

[‡] see Casella and Berger (2002) for details on the statistical concepts of consistency and efficiency.

[§] see Norden (1972) for historical survey.

3.2.1 Log-Likelihood functions

The log-likelihood functions for the univariate and bivariate Pearson IV distributions are given in the following two lemmas.

Lemma 3.3 (univariate log-likelihood function) *Let $\{x_1, \dots, x_k\}$ be a random sample from a $PIV_1(\mu, \sigma, \gamma, \nu)$ distribution, then the log-likelihood function is given by*

$$\ln \mathcal{L}(\ \mu, \sigma, \gamma, \nu \mid \{x_1, \dots, x_k\}\ ) =$$

$$-\frac{\nu+1}{2}\sum_{i=1}^{k} \ln\left(1 + \frac{\left(\frac{x_i-\mu}{\sigma}\right)^2}{\nu}\right) + \gamma \sum_{i=1}^{k} \tan^{-1}\left(\frac{x_i-\mu}{\sigma\sqrt{\nu}}\right) + k \ln\left(\frac{\mathcal{C}_1(\gamma,\nu)}{\sigma}\right) \tag{33}$$

Proof.

$$\begin{aligned} &\mathcal{L}(\mu, \sigma, \gamma, \nu \mid \{x_1, \dots, x_k\}) \\ &\qquad = \prod_{i=1}^{k} f(x_i \mid \mu, \sigma, \gamma, \nu) \\ &\qquad = \prod_{i=1}^{k} \underbrace{\mathcal{C}_1(\mu, \sigma, \gamma, \nu)}_{=\frac{\mathcal{C}_1(\gamma,\nu)}{\sigma}\text{, using Lemma 2.3.}} \left(1 + \frac{\left(\frac{x_i-\mu}{\sigma}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} e^{\gamma \tan^{-1}\left(\frac{x_i-\mu}{\sigma\sqrt{\nu}}\right)} \\ &\qquad \Rightarrow \quad (33). \end{aligned}$$

$\square$

Lemma 3.4 (bivariate log-likelihood function) *Let* $\left\{\begin{bmatrix}x_1\\y_1\end{bmatrix},\ldots,\begin{bmatrix}x_k\\y_k\end{bmatrix}\right\}$ *be a random sample from a* $PIV_2(\,\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\gamma_1,\gamma_2,\nu\,)$ *distribution, then the log-likelihood function is given by*

$$\ln\mathcal{L}\left(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\gamma_1,\gamma_2,\nu\;\middle|\;\left\{\begin{bmatrix}x_1\\y_1\end{bmatrix},\ldots,\begin{bmatrix}x_k\\y_k\end{bmatrix}\right\}\right)=$$
$$-\frac{\nu+2}{2}\sum_{i=1}^{k}\ln\left(1+\frac{\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_i-\mu_1}{\sigma_1}\right)\left(\frac{y_i-\mu_2}{\sigma_2}\right)+\left(\frac{y_i-\mu_2}{\sigma_2}\right)^2}{\nu\,(1-\rho^2)}\right)$$
$$+\gamma_1\sum_{i=1}^{k}\tan^{-1}\left(\frac{x_i-\mu_1}{\sigma_1\sqrt{\nu}}\right)+\gamma_2\sum_{i=1}^{k}\tan^{-1}\left(\frac{y_i-\mu_2}{\sigma_2\sqrt{\nu}}\right)+k\ln\left(\frac{\mathcal{C}_2(\rho,\gamma_1,\gamma_2,\nu)}{\sigma_1\sigma_2}\right) \tag{34}$$

Proof.

$$\mathcal{L}\left(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\gamma_1,\gamma_2,\nu\;\middle|\;\left\{\begin{bmatrix}x_1\\y_1\end{bmatrix},\ldots,\begin{bmatrix}x_k\\y_k\end{bmatrix}\right\}\right)$$
$$=\prod_{i=1}^{k} f(\,x_i,y_i\mid\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\gamma_1,\gamma_2,\nu\,)$$
$$=\prod_{i=1}^{k}\frac{\overbrace{\mathcal{C}_2(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\gamma_1,\gamma_2,\nu)}^{=\frac{\mathcal{C}_2(\rho,\gamma_1,\gamma_2,\nu)}{\sigma_1\sigma_2},\text{ using Lemma 2.7.}}\quad e^{\gamma_1\tan^{-1}\left(\frac{x-\mu_1}{\sigma_1\sqrt{\nu}}\right)+\gamma_2\tan^{-1}\left(\frac{y-\mu_2}{\sigma_2\sqrt{\nu}}\right)}}{\left(1+\frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}}$$

$$\Rightarrow\quad(34).$$

□

3.3 Multivariate Interpolation

During numerical maximum likelihood, values of the special functions $\mathcal{C}_1(\gamma, \nu)$ in the univariate case, and $\mathcal{C}_2(\rho, \gamma_1, \gamma_2, \nu)$ in the bivariate case, are needed at each step of the algorithm. Instead of repeatedly using numerical integration to obtain these values, I can precompute nodes on a dense regular grid at high precision and cache them in a multidimensional array. Intermediate values between the nodes can be estimated using *multivariate interpolation*. Both bi/tri/quad-linear interpolation and Catmull-Rom splines are considered, but only splines work properly.

3.3.1 Bilinear, Trilinear, and Quadlinear

In two dimensions, a function of two variables $p(x, y)$ can be interpolated inside four lattice points $p_{00}(x_0, y_0)$, $p_{01}(x_0, y_1)$, $p_{10}(x_1, y_0)$, and $p_{11}(x_1, y_1)$ as follows (Figure 18).

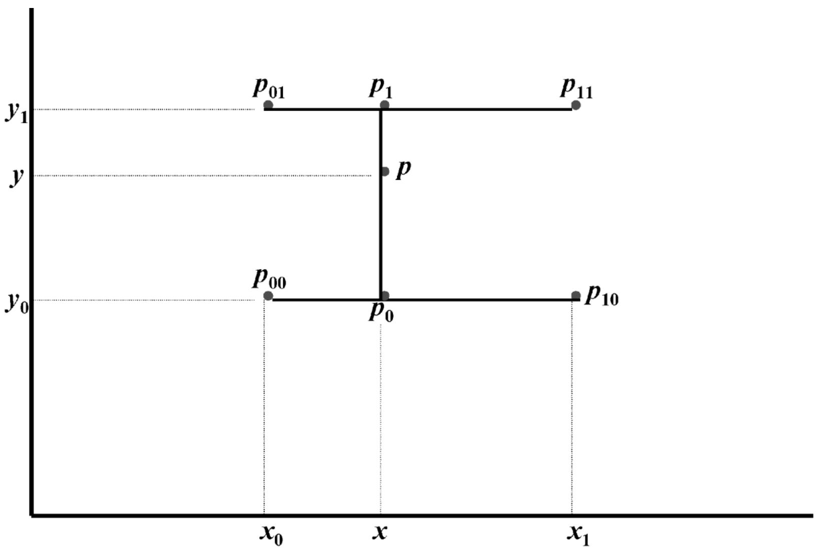


Figure 18: Bilinear Lattice (Kang, 1997)

To obtain value of point p in Figure 18, first hold y_0 constant and apply linear interpolation on lattice points p_{00} and p_{10} to obtain

$$p_0 = p_{00} + (p_{10} - p_{00})\frac{x - x_0}{x_1 - x_0}$$

Similarly,

$$p_1 = p_{01} + (p_{11} - p_{01})\frac{x - x_0}{x_1 - x_0}$$

by keeping y_1 constant. After obtaining p_0 and p_1, apply linear interpolation again keeping x constant

$$p(x, y) = p_0 + (p_1 - p_0)\frac{y - y_0}{y_1 - y_0}$$

A similar procedure can be used to interpolate a function inside a three-dimensional domain (Figure 19).

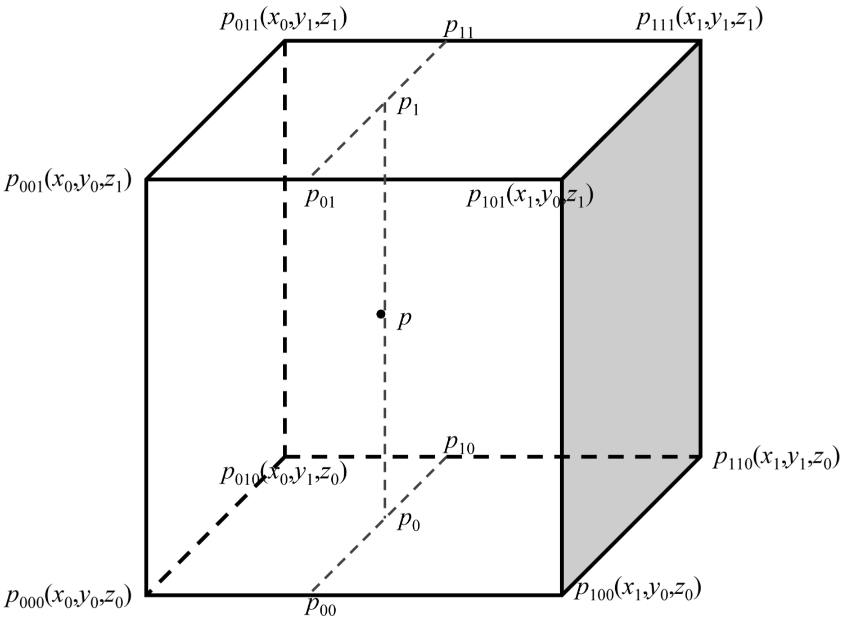


Figure 19: Trilinear Lattice (Kang, 1997)

Note that bilinear (Figure 20) and trilinear (Figure 21) interpolation surfaces are not planar, but nonlinear since they involve terms of order xy or xyz.

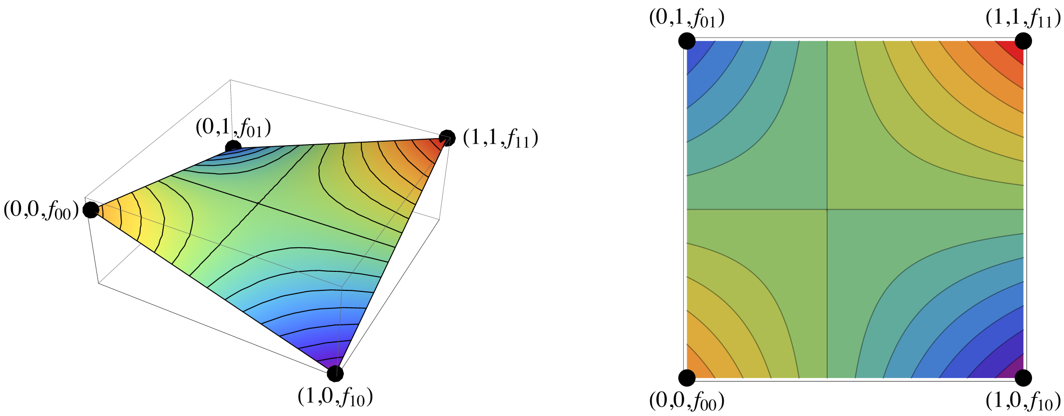


Figure 20: Bilinear Interpolation

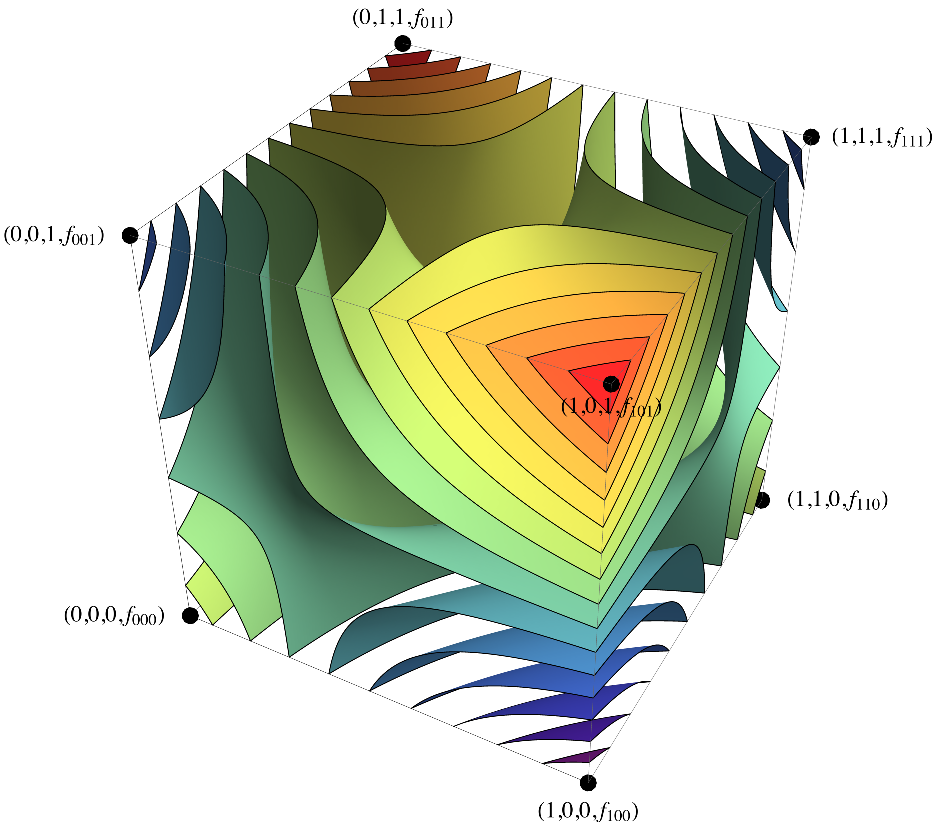


Figure 21: Trilinear Interpolation

3.3.2 Catmull-Rom Splines

By using Catmull-Rom splines (Catmull and Clark, 1978; Lalescu et al., 2009, 2010), the $\mathcal{C}_1(\gamma, \nu)$ special function can be interpolated with relative errors an order of magnitude better (Figure 22). Similar results were observed in higher dimensions.

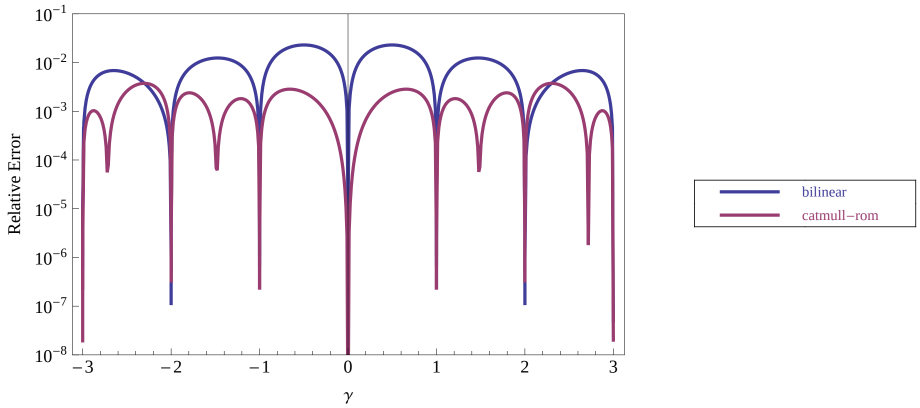


Figure 22: Relative Error - $\mathcal{C}_1(\gamma, 5)$

Bilinear interpolation has non-continuous derivatives near the nodes (Figure 23), creating false artifacts that numerical optimization wrongly converges to in numerical maximum likelihood fitting. Catmull-Rom spline interpolation, with C^0-continuous derivatives, avoids this problem.

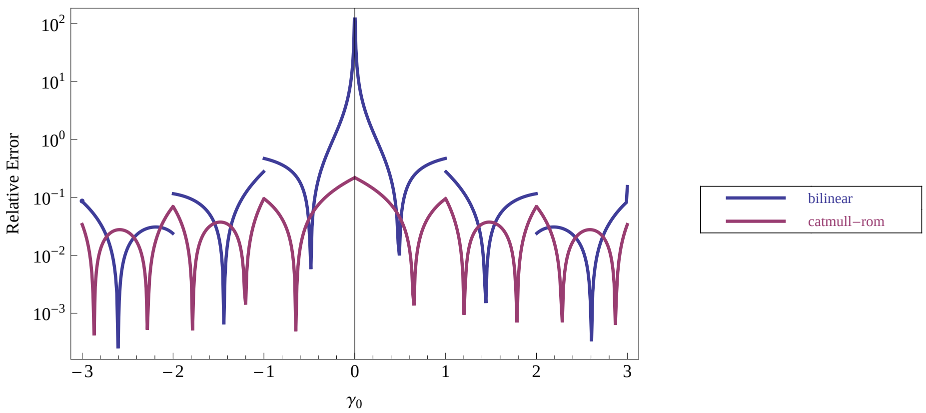


Figure 23: Relative Error - $\left.\frac{\partial}{\partial\gamma}\mathcal{C}_1(\gamma, \nu)\right|_{\gamma=\gamma_0, \nu=5}$

"The term Applied Mathematics encompasses many different areas of mathematical research with the driving motivation being a particular application or physical setting which researchers seek to understand more thoroughly. [...] The construction of a mathematical framework approximates (rarely can we be exact) the application of interest. This is as much an art as a science and will involve deciding upon which effects are important and which can be neglected."

- **David Amundsen** (http://math.carleton.ca/~dave/applied_math.htm).

4 Applications

4.1 Finance

Skewness and kurtosis have long been assumed to be negligible in many financial theories due to the complexity of handling them, but the Multivariate Pearson IV distribution proposed by this thesis can easily fit any empirical skewness, kurtosis, and correlation of financial portfolios.

4.1.1 Value-at-Risk (VaR)

Mainly due to new capital adequacy standards for banking and insurance, an increased interest exists in risk measures like Value-at-Risk (VaR). Bank regulators use VaR when determining the capital a bank is required to keep as a function of exposure to market risks (Jackson et al., 1998).

McNeil et al. (2005) defines VaR assuming an underlying probability distribution, which is referred to as *parametric VaR*.

Definition 4.1 (parametric VaR) *Given some confidence level $\alpha \in (0, 1)$ the VaR of a portfolio at the confidence level α is given by the smallest number x such that the probability that the loss L exceeds x is not larger than $(1 - \alpha)$*

$$VaR_\alpha = inf\{x \in \mathbb{R} : P(L > x) \leq 1 - \alpha\} = inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}.$$

Regardless of how VaR is computed, it should produce the correct number of breaks (within sampling error) in a backtest of the past. A common error is to report VaR based on the violated assumption that everything follows a multivariate normal distribution (Brown, 2002). Nassim Taleb calls the normal distribution the "Great Intellectual Fraud" in his widely-read book *The Black Swan* because people so often assume the normal distribution fits when it does not.

The financial events of the early 1990s found many firms in trouble because the same underlying bet had been made at many places in the firm in non-obvious ways, eg. the well-known *carry trade* which exploits imbalances between interest rates paid in different currencies. Since many trading desks already computed VaR and it was a common risk measure defined for all businesses that could be aggregated, it was the natural choice for reporting firmwide risk.

Development was most extensive at J. P. Morgan, which published the methodology and gave free access to estimates of the necessary underlying parameters in 1994.

In 1997, the U.S. Securities and Exchange Commission ruled that public corporations must disclose quantitative information about their derivatives activity. Major banks and dealers chose to implement the rule by including VaR information in the notes to their financial statements (Jorion, 2007). Worldwide adoption of the Basel II Accord, beginning in 1999, further propagated the use of VaR as the preferred measure of market risk.

J. P. Morgan CEO Dennis Weatherstone famously called for a “4:15 report” that combined all risk from his global financial enterprise on a single page, available within 15 minutes of the 4:00 market close in New York (Kolman et al., 1997). This clearly illustrates a requirement for computational efficiency of VaR statistical calculations, even for a major bank with an almost unlimited computer hardware budget.

Analytical expressions for the quantiles of the normal and t distributions are known. For the Pearson IV distribution, a given quantile α can be obtained numerically from the cdf by solving

$$\frac{\mathcal{C}_1(\gamma,\nu)\pi\sqrt{\nu}}{e^{-\frac{\gamma\pi}{2}}}\int_0^{\frac{1}{2}+\frac{\tan^{-1}\left(\frac{y-\mu}{\sigma\sqrt{\nu}}\right)}{\pi}} e^{\gamma\pi\theta}\sin^{\nu-1}(\pi\theta)\,d\theta=\alpha$$

for y.

The Cornish-Fisher expansion (Cornish and Fisher, 1938; Fisher and Cornish, 1960; Finney, 1963) provides a relationship between the moments of a distribution and its percentiles, which can also be used to calculate Pearson IV quantiles (Bowman and Shenton, 1979a,b; Davenport and Herring, 1979).

Example 4.2 (one-stock VaR) *Backtesting of 1-day 99% and 95% VaR, using stocks from Table 2, should result in* $\hat{\beta} = 1$ *and* $\hat{\beta} = 5$ *breaks*¶ *respectively for a 100-day period. Pearson IV performs similarly to normal and t distributions for* $\alpha = .01$*, but not for* $\alpha = .05$ *(Table 7).*

	$N(\mu, \sigma)$	$t(\mu, \sigma, \nu)$	$PIV(\mu, \sigma, \gamma, \nu)$
1-day 99% VaR			
Mean($\hat{\beta}$)	1.6359	1.1206	1.1158
Type I Error ($\hat{\beta} < 1$)	41.61%	48.46%	49.41%
Correct ($\hat{\beta} = 1$)	21.51%	24.59%	24.35%
Type II Error ($\hat{\beta} > 1$)	36.88%	26.95%	26.24%
1-day 95% VaR			
Mean($\hat{\beta}$)	5.0260	5.9740	7.6903
Type I Error ($\hat{\beta} < 5$)	54.85%	46.34%	35.46%
Correct ($\hat{\beta} = 5$)	8.98%	8.51%	7.57%
Type II Error ($\hat{\beta} > 5$)	36.17%	45.15%	56.97%

Table 7: VaR backtesting

¶also known as *failures* or *exceptions* in the literature.

4.1.2 Option Pricing using univariate Pearson IV

Few academic articles can claim to have created a trillion dollar industry. Black and Scholes (1973), together with Bachelier (1900), Merton (1973) and Ito (1944, 1946), can make such a claim. Shortly after these publications, stock options and other financial derivatives started trading publicly. People now had a formula to price these securities which until then were traded *over-the-counter (OTC)* between institutions at a microscopic volume. The *Chicago Board Options Exchange (CBOE)*, established in 1973, traded the first exchange-listed stock options. The CBOE remains one of the world's largest options exchanges.

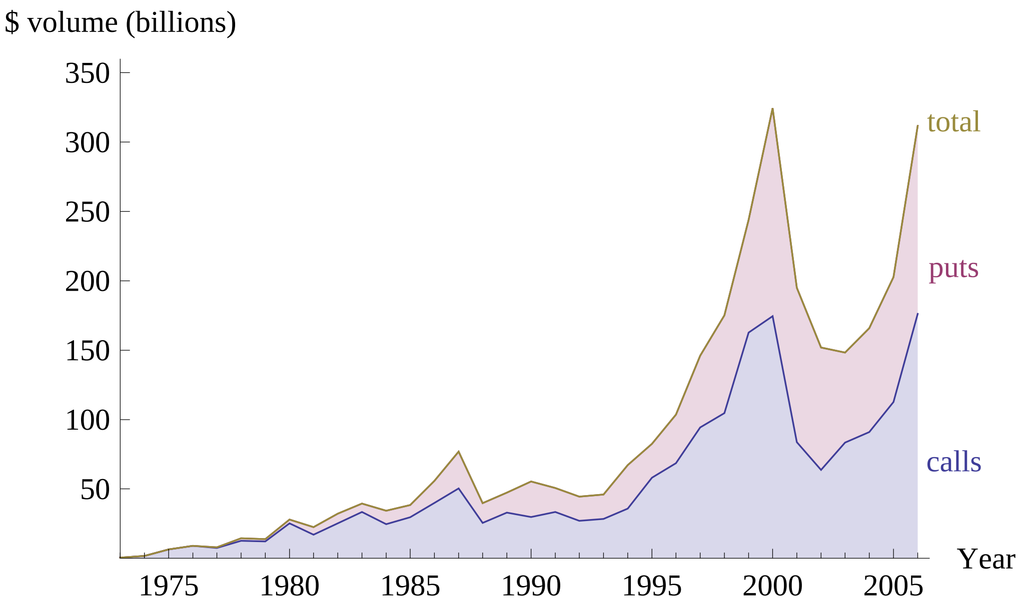


Figure 24: CBOE Dollar Volume 1973-2006

Following Black and Scholes' (1973) dynamical hedging PDE approach, Cox and Ross (1976) offered a simpler interpretation of the theoretical value of an option as the expected value of its payoff at expiry

$$C(r,K,T) = e^{-rT}\int_{-\infty}^{\infty} \max(s-K,0)\, w_T(s)\, ds = e^{-rT}\int_{K}^{\infty} (s-K)\, w_T(s)\, ds \qquad (35)$$

$$P(r,K,T) = e^{-rT}\int_{-\infty}^{\infty} \max(K-s,0)\, w_T(s)\, ds = e^{-rT}\int_{-\infty}^{K} (K-s)\, w_T(s)\, ds \qquad (36)$$

where r is the riskless interest rate, K is the strike price, T is the time to expiry, and e^{-rT} is the time-value-of-money discounting factor.

The Black-Scholes option formula (77) is recovered from (35) or (36) when a lognormal distribution is used for $w_T(s)$. If there exists a more accurate model of stock prices than geometric brownian motion, and a corresponding $w_T(s)$ distribution at expiry, then better estimates of the actual value of a call or put option become available.

Stock options are the mechanism with which one can profit from better estimates of the underlying probability distribution of stock market prices. There are no *free lunches*, only *cheaper lunches*. Some might sell their lunch for less than its value, others might be willing to buy yours for more than its worth. Profitable trading opportunities are created by deviations away from rational pricing, best explained by a relatively new field known as *behavioral finance*.[‖]

[‖]see §A.2 for literature survey.

Theorem 4.3 (option profit distribution) *Let s be the strike price of an option, and a be the asking price. If the log returns $\{\ln P(t) - \ln P(t-d)\}$ of the underlying stock are distributed as a random variable $X \sim f_X$, then the random variable Y representing the profit at expiry occurring at time $t+d$ is:*

$$CALL\ PROFIT\ \% \sim f_Y(y) = \frac{a\ f_X\big(\ln(a(y+1)+s) - \ln P(t)\big)}{a(y+1)+s}$$

$$PUT\ PROFIT\ \% \sim f_Y(y) = \frac{a\ f_X\big(\ln(s-a(y+1)) - \ln P(t)\big)}{s-a(y+1)}$$

Proof.

$$\text{CALL PROFIT \%} = \frac{payout - a}{a} = \frac{P(t+d) - s - a}{a} = \frac{P(t)e^X - s - a}{a} = g(X)$$

$$\Rightarrow \quad P(t)e^X = aY + s + a = a(Y+1) + s$$

$$\Rightarrow \quad e^X = \frac{a(Y+1)+s}{P(t)}$$

$$\Rightarrow \quad X = \ln(a(Y+1)+s) - \ln P(t) = g^{-1}(Y)$$

$$\therefore \quad f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|$$

$$= \frac{a\ f_X\big(\ln(a(y+1)+s) - \ln P(t)\big)}{a(y+1)+s}$$

and similarly for put options

$$\text{PUT PROFIT \%} = \frac{payout - a}{a} = \frac{s - P(t+d) - a}{a} = \frac{s - P(t)e^X - a}{a} = g(X)$$

$$\Rightarrow \quad aY = s - P(t)e^X - a = a(Y+1) + s$$

$$\Rightarrow \quad P(t)e^X = s - aY - a = s - a(Y+1)$$

$$\Rightarrow \quad e^X = s - aY - a = \frac{s - a(Y+1)}{P(t)}$$

$$\Rightarrow \quad X = \ln\left(s - a(Y+1)\right) - \ln P(t) = g^{-1}(Y)$$

$$\begin{aligned} \therefore \quad f_Y(y) &= f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right| \\ &= \frac{a\ f_X\Big(\ln\left(s - a(y+1)\right) - \ln P(t)\Big)}{s - a(y+1)} \end{aligned}$$

$\square$

Based on an option profit distribution from Theorem 4.3, various metrics to select favorable option positions are given by Definition 4.4.

Definition 4.4 (option odds, expected profit, and risk ratio) *For any option, the odds* (Θ)*, ie. probability that an option trade will be profitable, are defined as*

$$\Theta = \int_0^\infty f(x)dx \tag{37}$$

and the expected profit is

$$\mathbb{E}(\mathit{profit}) = \int_{-\infty}^\infty x f(x)dx \tag{38}$$

and the Psi risk ratio is

$$\Psi = \frac{\int_0^\infty x f(x)dx}{\int_{-\infty}^0 x^2 f(x)dx} \tag{39}$$

Figures 25, 26, and 27 show trading simulation results for the 2002-2005 period, with options being selected by best odds, expected profit, and Psi risk ratio. Expected profit is clearly the best selection method earning an average 47.18% return per year, while odds and psi risk ratio yield disappointing results. According to a Wilcoxon signed-rank test (Wilcoxon, 1945; Siegel, 1956) on net monthly gains for the expected profit strategy, PIV$(\mu, \sigma, \gamma, \nu)$ outperforms both the N(μ, σ) and $t(\mu, \sigma, \nu)$ distributions with $p = 0.000249$ and $p = 0.000039$ respectively, where the p-values are obtained using the moment generating method of Mitic (1996).

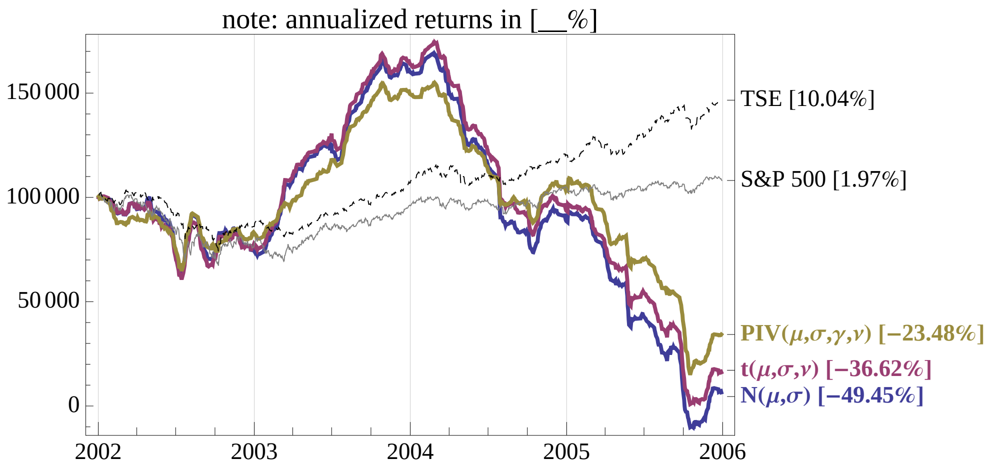


Figure 25: Option Trading Simulation - by Odds

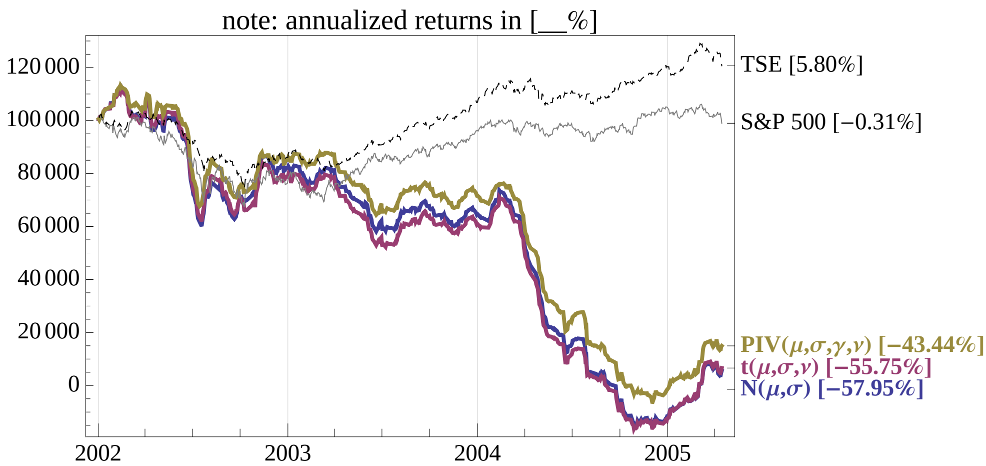


Figure 26: Option Trading Simulation - by Psi Risk Ratio

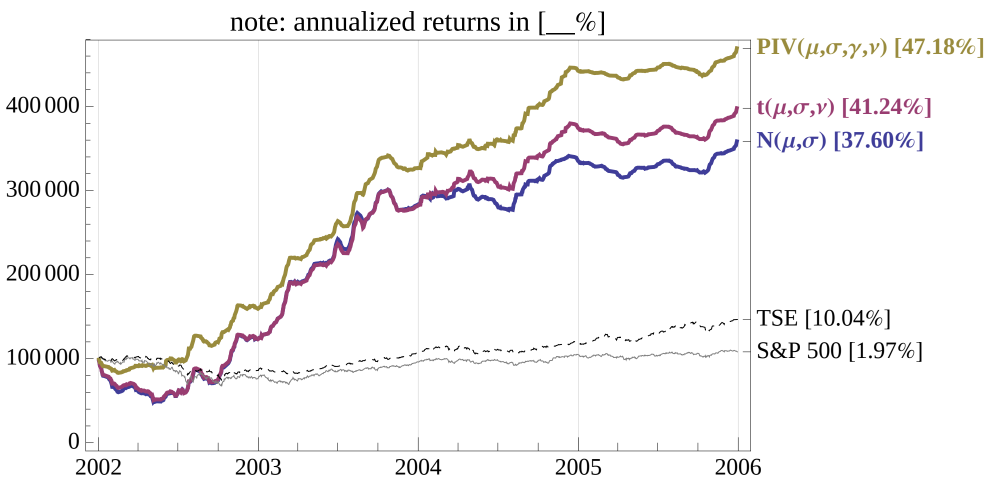


Figure 27: Option Trading Simulation - by Expected Profit

4.1.3 Option Portfolio Pricing using multivariate Pearson IV

While there is considerable literature on the univariate *option pricing problem*, much less has been published on the *option portfolio pricing problem* (Merton et al., 1978; Bookstaber and Clarke, 1984). By assuming log returns follow an MPIV distribution and using the following Theorem, the expectation of any function of prices at expiry can be computed as illustrated by Example 4.6 in the bivariate case.

Theorem 4.5 (expected value of function of stock prices at expiry) *Let an option pair on stocks X and Y, currently trading at prices P_1 and P_2 respectively, have joint log returns distributed as a $BPIV(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2, \nu)$ random vector. Then, for given prices at expiry U and V, the expected value of any function $h(U, V)$ is*

$$E\Big(h(u,v)\big|\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\gamma_1,\gamma_2,\nu\Big) = \mathcal{C}_2(\rho,\gamma_1,\gamma_2,\nu)\,\pi^2\nu\, e^{\frac{\pi}{2}(\gamma_1+\gamma_2)}$$

$$\times \int_0^1\int_0^1 \frac{h\left(P_1 e^{\mu_1-\sigma_1\sqrt{\nu}\cot(\pi\theta)}, P_2 e^{\mu_2-\sigma_2\sqrt{\nu}\cot(\pi\phi)}\right) e^{\pi(\gamma_1\theta+\gamma_2\phi)}\; d\theta d\phi}{\sin^2(\pi\theta)\sin^2(\pi\phi)\left(1+\dfrac{\cot^2(\pi\theta)-2\rho\cot(\pi\theta)\cot(\pi\phi)+\cot^2(\pi\phi)}{(1-\rho^2)}\right)^{\frac{\nu+2}{2}}} \tag{40}$$

Proof. From the random variables X and Y as the joint distribution of log returns, transform to (U, V) as the joint distribution of prices at expiry

$$U = P_1 e^X, \qquad V = P_2 e^Y$$

$$\Rightarrow \quad x = \ln u - \ln P_1 = g_1^{-1}(u,v), \qquad y = \ln v - \ln P_2 = g_2^{-1}(u,v).$$

$$\Rightarrow \quad \frac{\partial x}{\partial u} = \frac{1}{u}, \qquad \frac{\partial x}{\partial v} = 0, \qquad \frac{\partial y}{\partial u} = 0, \qquad \frac{\partial y}{\partial v} = \frac{1}{v}.$$

$$\Rightarrow \quad f_{U,V}(u,v) = f_{X,Y}\left(g_1^{-1}(u,v), g_2^{-1}(u,v)\right)|J|$$

$$= \frac{f_{X,Y}\left(\ln u - \ln P_1, \ln v - \ln P_2\right)}{u\,v}, \quad 0 \le u, v < \infty$$

$$= \frac{\left(\frac{\mathcal{C}_2(\rho,\gamma_1,\gamma_2,\nu)}{\sigma_1\,\sigma_2\,u\,v}\right)\; e^{\gamma_1 \tan^{-1}\left(\frac{\ln u - \ln P_1 - \mu_1}{\sigma_1\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{\ln v - \ln P_2 - \mu_2}{\sigma_2\sqrt{\nu}}\right)}}{\left(1 + \frac{\left(\frac{\ln u - \ln P_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\ln u - \ln P_1 - \mu_1}{\sigma_1}\right)\left(\frac{\ln v - \ln P_2 - \mu_2}{\sigma_2}\right) + \left(\frac{\ln v - \ln P_2 - \mu_2}{\sigma_2}\right)^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}}$$

and therefore the expected value of $h(u,v)$ is

$$E\Big(h(u,v)\Big|\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\gamma_1,\gamma_2,\nu\Big) =$$

$$\int_0^\infty\int_0^\infty \frac{h(u,v)\left(\frac{\mathcal{C}_2(\rho,\gamma_1,\gamma_2,\nu)}{\sigma_1\,\sigma_2\,u\,v}\right)\; e^{\gamma_1 \tan^{-1}\left(\frac{\ln u - \ln P_1 - \mu_1}{\sigma_1\sqrt{\nu}}\right) + \gamma_2 \tan^{-1}\left(\frac{\ln v - \ln P_2 - \mu_2}{\sigma_2\sqrt{\nu}}\right)}}{\left(1 + \frac{\left(\frac{\ln u - \ln P_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\ln u - \ln P_1 - \mu_1}{\sigma_1}\right)\left(\frac{\ln v - \ln P_2 - \mu_2}{\sigma_2}\right) + \left(\frac{\ln v - \ln P_2 - \mu_2}{\sigma_2}\right)^2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}}\,du\,dv$$

Substitute

$$\theta = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{\ln u - \ln P_1 - \mu_1}{\sigma_1\sqrt{\nu}}\right), \quad \phi = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{\ln v - \ln P_2 - \mu_2}{\sigma_2\sqrt{\nu}}\right)$$

$$\Rightarrow \quad \frac{\ln u - \ln P_1 - \mu_1}{\sigma_1\sqrt{\nu}} = -\cot(\pi\theta) \quad \Rightarrow \quad u = P_1 e^{\mu_1 - \sigma_1\sqrt{\nu}\cot(\pi\theta)}$$

$$\Rightarrow du = \left(P_1 e^{\mu_1 - \sigma_1\sqrt{\nu}\cot(\pi\theta)}\right)\left(\pi\sigma_1\sqrt{\nu}\csc^2(\pi\theta)\right)d\theta = u\pi\sigma_1\sqrt{\nu}\csc^2(\pi\theta)d\theta$$

and similarly

$$v = P_2 e^{\mu_2 - \sigma_2\sqrt{\nu}\cot(\pi\phi)} \quad \Rightarrow \quad dv = v\pi\sigma_2\sqrt{\nu}\csc^2(\pi\phi)d\phi$$

Next as $u \to 0$, $\tan^{-1}\left(\frac{\ln u - \ln P_1 - \mu_1}{\sigma_1\sqrt{\nu}}\right) \to -\frac{\pi}{2}$, so $\theta \to \frac{1}{2} + \frac{-\cancel{\pi}/2}{\cancel{\pi}} = 0$ and as $u \to \infty$, $\theta \to \frac{1}{2} + \frac{\cancel{\pi}/2}{\cancel{\pi}} = 1$. Similarly as $v \to 0, \phi \to 0$, and as $y \to \infty, \phi \to 1$. Therefore

$$E\Big(h(u,v)\Big|\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\gamma_1,\gamma_2,\nu\Big) =$$

$$\int_0^1\int_0^1 h(u,v)\left(\frac{\mathcal{C}_2(\rho,\gamma_1,\gamma_2,\nu)}{\cancel{\sigma_1}\cancel{\sigma_2}\cancel{u}\cancel{v}}\right)\ e^{\gamma_1\left(\pi(\theta-\frac{1}{2})\right)+\gamma_2\tan^{-1}\left(\pi(\phi-\frac{1}{2})\right)}$$

$$\times \frac{\left(\cancel{u}\pi\cancel{\sigma_1}\sqrt{\nu}\csc^2(\pi\theta)d\theta\right)\left(\cancel{v}\pi\cancel{\sigma_2}\sqrt{\nu}\csc^2(\pi\phi)d\phi\right)}{\left(1+\frac{\cancel{\nu}\cot^2(\pi\theta)-2\rho\cancel{\sqrt{\nu}}\cot(\pi\theta)\cancel{\sqrt{\nu}}\cot(\pi\phi)+\cancel{\nu}\cot^2(\pi\phi)}{\cancel{\nu}(1-\rho^2)}\right)^{\frac{\nu+2}{2}}}$$

$$= (40)$$

$\square$

The following example shows how to apply Theorem 4.5 to obtain the bivariate odds and expected profit for a option portfolio consisting of two options, each of which can be a CALL or a PUT. The indicator function $i(u,v)$ and payout function $p(u,v)$ are non-zero in integration regions which vary depending on whether an option pair is CALL/CALL, CALL/PUT, PUT/CALL, or PUT/PUT (Figure 28).

Example 4.6 (bivariate odds and expected profit) *To calculate odds for a given option pair* $\{CALL(s_1,a_1), PUT(s_2,a_2)\}$*, use Theorem 4.5 with the indicator function*

$$i(u,v) = \begin{cases} 1, & (u-s_1)^+ + (s_2-v)^+ > a_1 + a_2 \\ 0, & otherwise \end{cases}$$

where $(x)^+ = \max(x,0)$*,* s_1, s_2 *are the strike prices, and* a_1, a_2 *the ask prices. Similarly, to calculate expected profit for the same option pair use the payout function*

$$p(u,v) = (u-s_1)^+ + (s_2-v)^+ .$$

This example can easily be generalized to a portfolio with an arbitrary number of options, however the integration region becomes rather complicated even for the trivariate case (Figure 29).

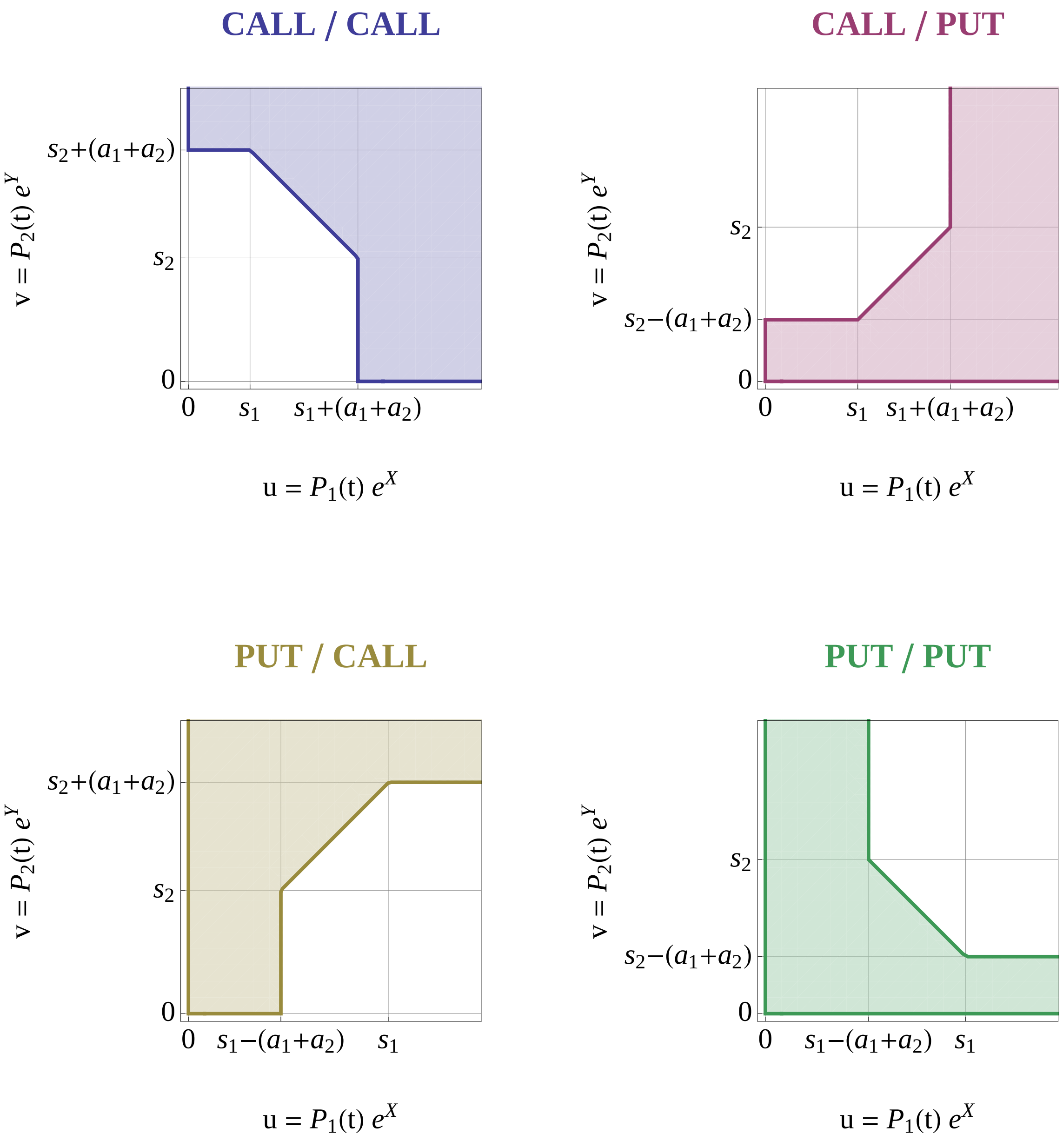


Figure 28: Bivariate Option Portfolio - Integration Regions

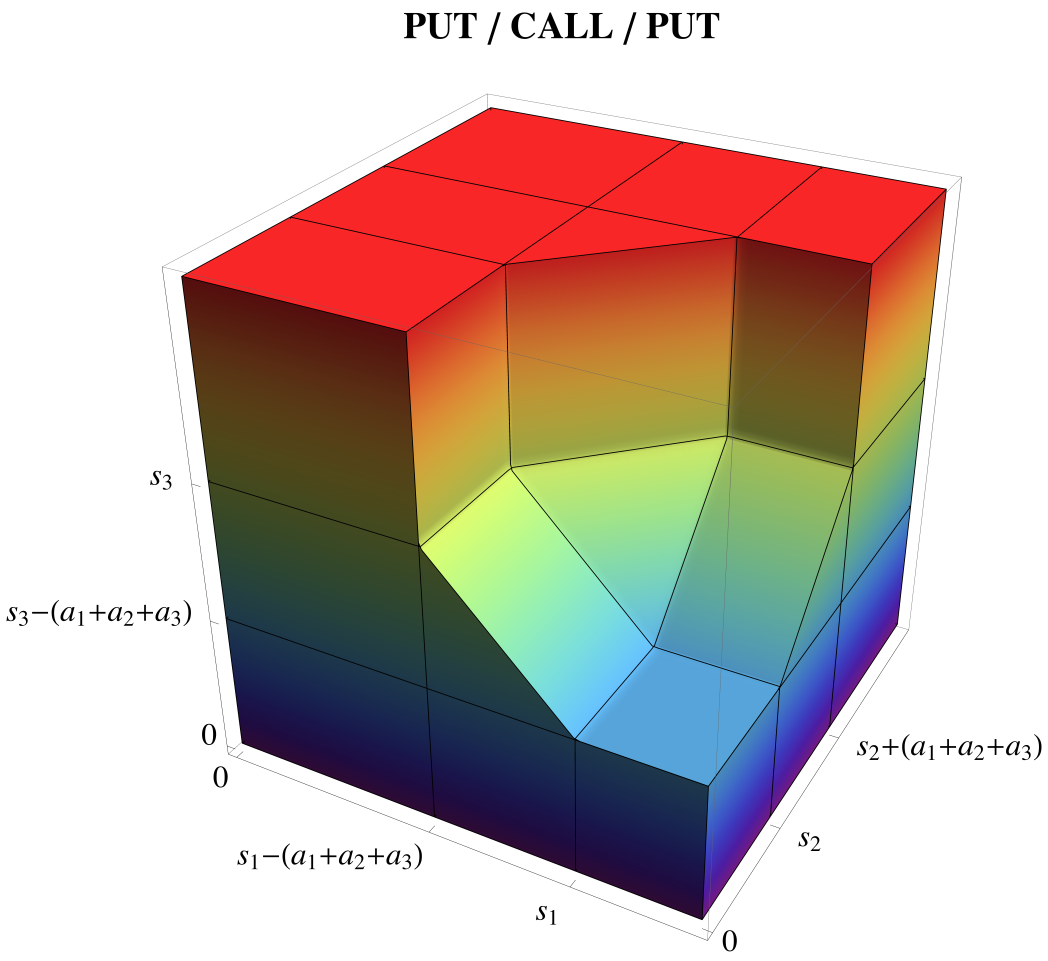


Figure 29: Trivariate Option Portfolio - Integration Region

4.2 Climate Change

In the study of climate change, one of the most pressing issues is to quantify the uncertainty around greenhouse gas emissions estimates. National emissions inventories are prepared using various methodologies depending on available information. As such, emission estimates are wide-ranging in their accuracy and associated uncertainty. The scientific community must make every effort to present an unbiased assessment of point estimates and confidence intervals conditioned on methodologies and data available at the time of publication.

Currently, few countries are conducting sophisticated uncertainty analyses using Monte Carlo numerical simulation. As regulatory requirements and economics of carbon trading markets mature, we will need to capture model complexities beyond the capabilities of the normal distribution. The Pearson IV distribution, with its ability to fit any empirical skewness and kurtosis, will eventually be needed in applied mathematical models of biological and industrial processes.

Historically, the most common approach to uncertainty in government policy analysis has been to ignore it. Why consider uncertainty?

- Uncertainty information is primarily intended to help prioritize efforts in improving the accuracy of inventories, and guide decisions on methodological choice.
- The greater the uncertainty, the greater the expected value of additional information.
- Even when there seems to be virtually no empirical information about the likely value of a quantity, there is often good evidence to provide bounds on it.

4.2.1 IPCC Tier 2 - Monte Carlo Uncertainty Analysis

Murphy et al. (2009) perform an uncertainty analysis of emission/removals estimates from agriculture and land-use, land-use change and forestry (LULUCF). This uncertainty analysis quantifies the level of confidence in the emissions estimates, by generating error bounds around each variable in the methodology used to estimate emissions. Uncertainties arising from both activity data and emission factors are combined to obtain an overall measure of uncertainty for total emissions (Figure 30).

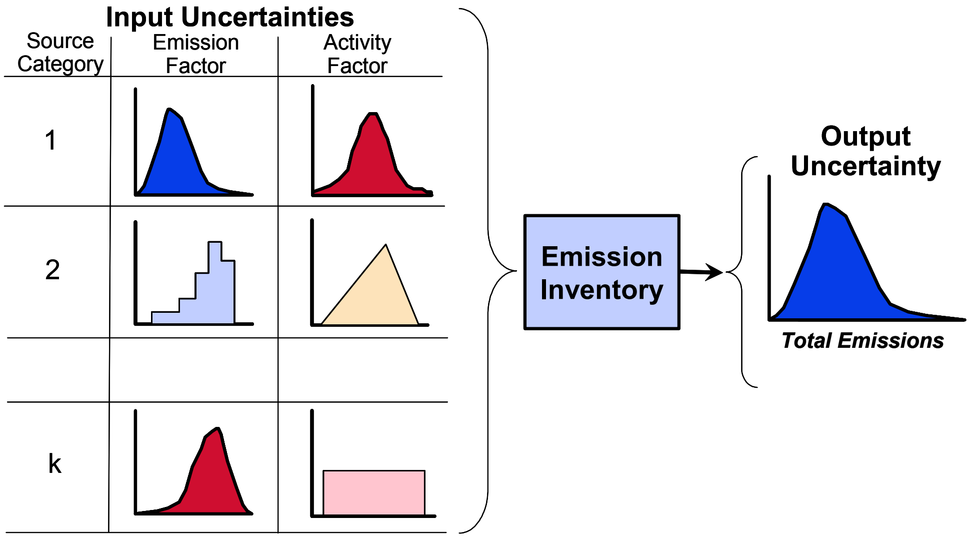


Figure 30: Input and Output Uncertainty (Frey, 2007)

Under the *United Nations Framework Convention on Climate Change* (UNFCCC), Annex I Parties are required to quantitatively estimate the uncertainties of their National Greenhouse Gas Inventories using good practice guidance (Frey et al., 2006) from the *Intergovernmental Panel on Climate Change* (IPCC). Environment Canada has quantitied the uncertainty of its GHG estimates for many years (McCann, 1994; SGA, 2000; ICF, 2004, 2005), other countries have also (Winiwarter and Rypdal, 2001; Webster et al., 2003; Davies et al., 2006; Gosling and O'Hagan, 2007; Kennedy et al., 2008). Murphy et al. (2009) follow a hybrid IPCC *Tier 1 (Analytical)* and *Tier 2 (Monte Carlo)* method of uncertainty analysis, also discussed in NARSTO (2005), which involves these steps:

- *Document the methodology* from a top-down perspective, ie. how are estimates calculated (Figures 31 and 32);
- *Replicate the methodology* and confirm estimates are the same as previously obtained;
- *Figure out the inputs* to the model, and their uncertainty;
- *Propagate the errors* from inputs through the methodology to the estimates (either using tier 1 or 2);
- *Correlated errors* arise when an input is used in more than one subcategory of the model.

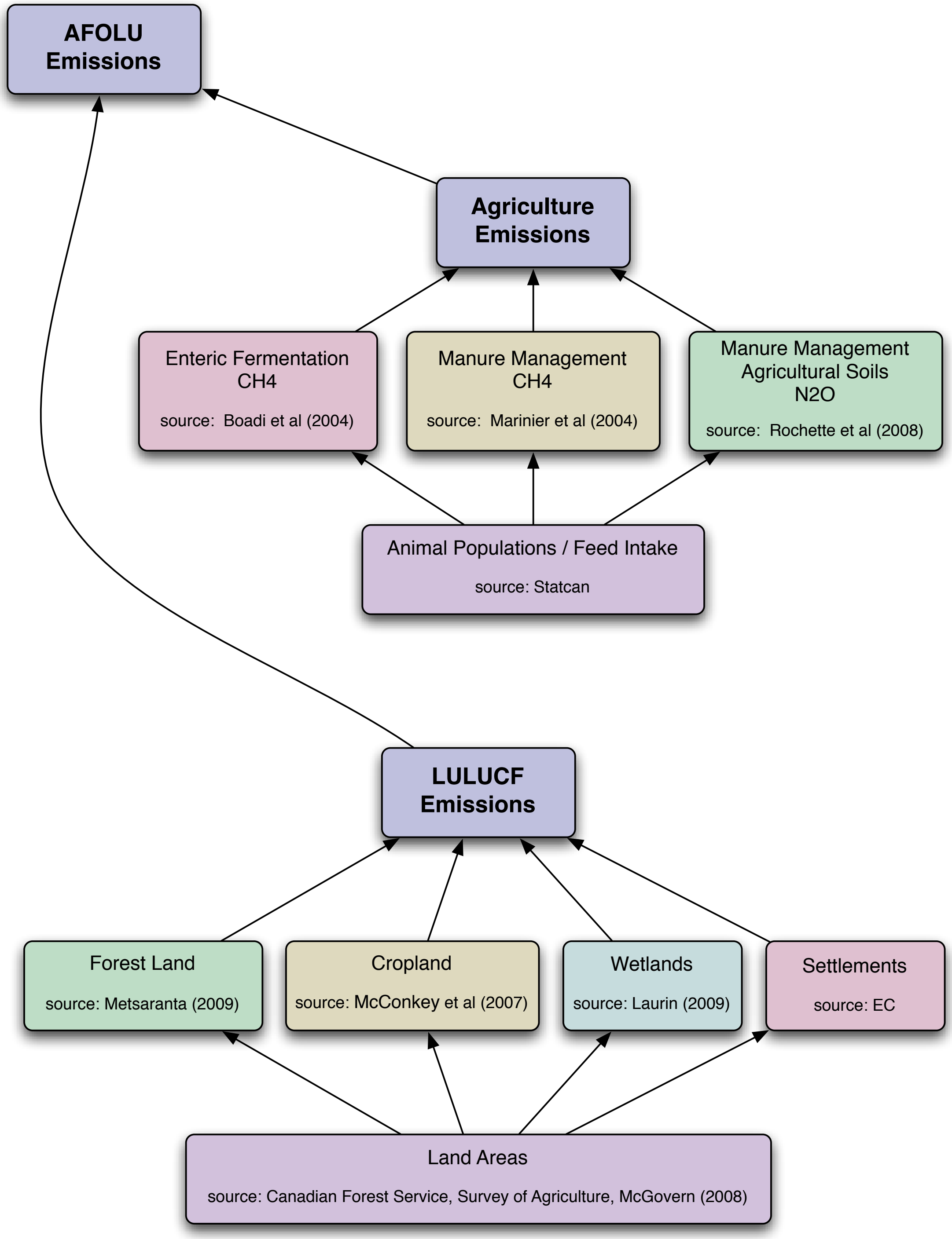


Figure 31: Model Diagram - AFOLU (Murphy et al., 2009)

In the forest land sector, emission factors usually have a long positive or negative tail (as the case may be for either an emission factor or a removal factor), and so we need to allow for a certain degree of skewness, which is currently modeled using the lognormal or Weibull distributions. The general model (Figure 32) for uncertainty of all forest land related terms in year i is

$$\begin{aligned} FL(i) &= FLFL(i) + FLL(i) + LFL(i) \\ &= A_{FLFL}(i) \cdot EF_{FLFL}(i) + A_{FLL}(i) \cdot EF_{FLL}(i) + A_{LFL}(i) \cdot EF_{LFL}(i) \end{aligned} \tag{41}$$

where

$FLFL(i)$:	emissions from forest land remaining forest land in year i;
$A_{FLFL}(i)$:	area of forest remaining forest in year i;
$EF_{FLFL}(i)$:	emission factor for forest remaining forest in year i;
$FLL(i)$:	emissions from deforestation in year i;
$A_{FLL}(i)$:	area deforested in year i, ie. $FLL = FLCL + FLWL + FLSL$;
$EF_{FLL}(i)$:	emission factor for deforestation in year i;
$LFL(i)$:	emissions from afforestation in year i;
$A_{LFL}(i)$:	area afforested in year i, ie. L-FL;
$EF_{LFL}(i)$:	emission factor for afforestation in year i.

Both *Net Biome Production* and *Net Ecosystem Production* are the result of one lognormal random variable sustracted from another lognormal random variable, yielding a skewed quantity with infinite support which is best modeled using a Pearson IV random variable. This is the subject of the following example.

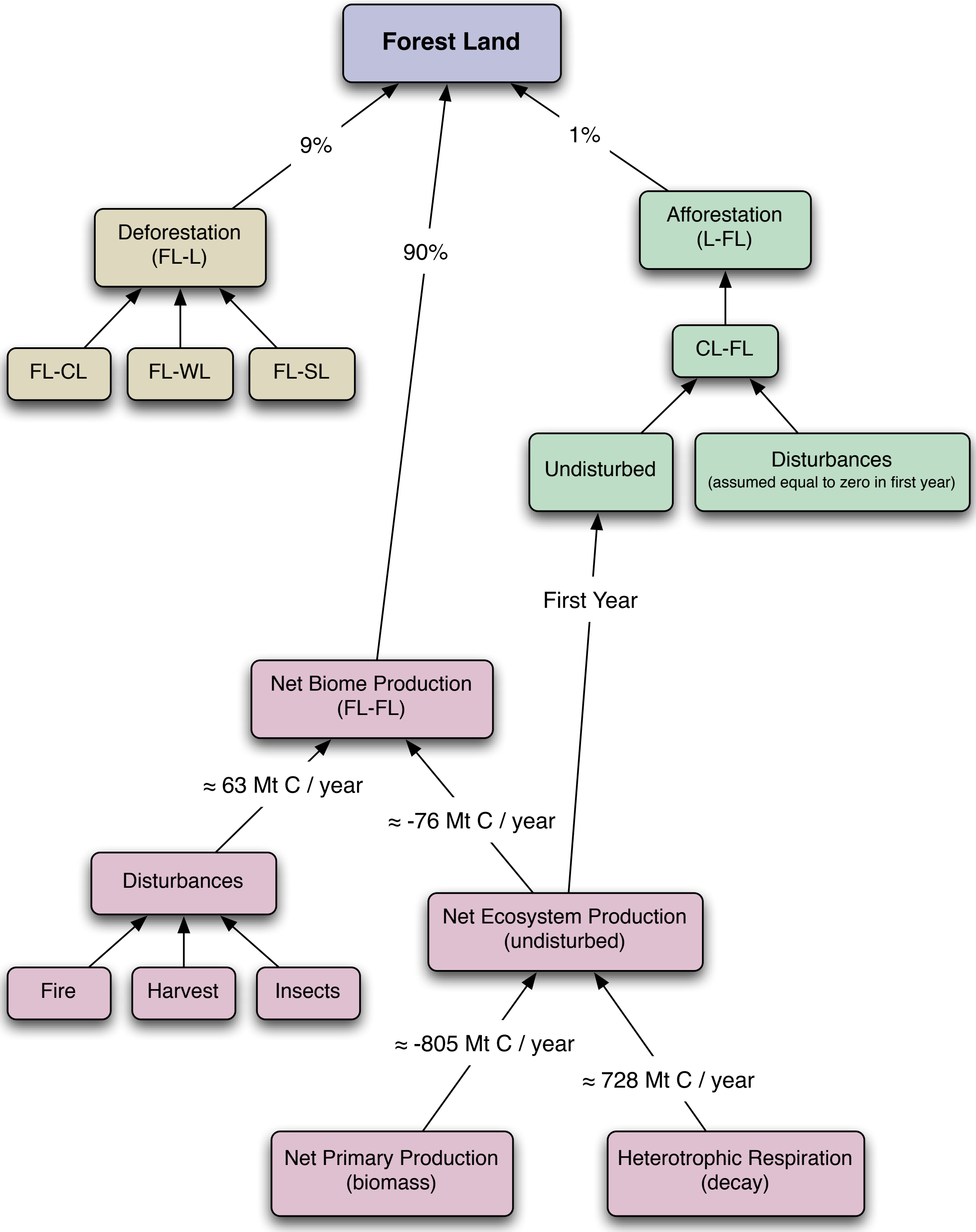


Figure 32: Model Diagram - Forest Land (Murphy et al., 2009)

Example 4.7 (LogNormal(μ_1, σ_1) - LogNormal(μ_2, σ_2)) *In Monte Carlo simulation, the distribution of $X-Y$, where $X \sim LogNormal(\mu_1, \sigma_1)$ and $Y \sim LogNormal(\mu_2, \sigma_2)$ are both lognormal random variables, can be represented by a Pearson IV random variable, ie.*

$$LogNormal(\mu_1, \sigma_1) - LogNormal(\mu_2, \sigma_2) \sim PIV(\mu_3, \sigma_3, \gamma, \nu)$$

If X and Y both have $[0, \infty)$ support, then $-Y$ has $(-\infty, 0]$ support. Hence, $X - Y$ has support on $\mathbb{R}$, and cannot be represented by a lognormal random variable. Weibull distributions are commonly used to represent skewness in the exponential family, but the Weibull distribution is also restricted to $[0, \infty)$ support. $X-Y$ can be approximated using a normal distribution with support on $\mathbb{R}$, by ignoring the non-normal skewness and kurtosis. With support on $\mathbb{R}$ and the ability to represent non-normal skewness and kurtosis, a Pearson IV distribution is best suited for the random variable $X - Y$.

4.3 Pearson Diffusions

A *Pearson diffusion* (Forman and Sørensen, 2008; Sørensen, 2009) is a stationary solution to a stochastic differential equation (SDE) of the form

$$dX(t) = -\theta\left(X(t)-\mu\right)dt + \sqrt{2\theta\left(aX(t)^2+bX(t)+c\right)}dW(t) \qquad (42)$$

where $\theta > 0$ is a scaling of time that determines how fast the diffusion moves, and $W(t)$ is a Brownian motion or Wiener process (Definition A.19). The parameters μ, a, b and c determine the state space of the diffusion as well as the shape of the invariant distribution, eg. μ is the mean of the invariant distribution.

The scale and speed densities of diffusion (42) are

$$s(x) = e^{\int_{x_0}^{x} \frac{y-\mu}{ay^2+by+c}dy}, \qquad m(x) = \frac{1}{s(x)(ax^2+bx+c)}$$

and the invariant distribution has density proportional to the speed density

$$\frac{dm(x)}{dx} = -\frac{(2a+1)x-\mu+b}{ax^2+bx+c}m(x)$$

which implies that when a stationary solution to (42) exists, the invariant distribution belongs to the Pearson system. For example, when $\sigma^2(x) = 2\theta$ there exists a unique *Ornstein-Uhlenbeck process* solution to (42) where the invariant distribution is the normal distribution with mean 0 and variance 1.

Figure 33 shows numerical solutions for the linear SDDE** family

$$dX(t) = \big(aX(t) + bX(t-r)\big)dt + \sigma dW(t)$$

where r stretches solutions horizontally, and σ stretches them vertically. Both r and σ can be normalized to 1 for simplification purposes while retaining the qualitative features of solutions. The stability region for parameters a and b is shaded in blue.

Consistent with results from Küchler and Platen (2007), I was not able to replicate non-normal skewness and kurtosis in my numerical experiments with linear SDDEs, unless the diffusion term also had a delay term, ie.

$$dX(t) = \big(\alpha X(t) + \beta X(t-1)\big)dt + \big(\lambda X(t) + \delta X(t-1)\big)dW(t)$$

$$= \begin{bmatrix} \alpha X(t) + \beta X(t-1) & \lambda X(t) + \delta X(t-1) \end{bmatrix} \begin{bmatrix} dt \\ dW(t) \end{bmatrix}$$

$$= \begin{bmatrix} X(t) & X(t-1) \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \lambda & \delta \end{bmatrix} \begin{bmatrix} dt \\ dW(t) \end{bmatrix}$$

where $\delta \neq 0$. These findings would be consistent with an hypothesis that non-normal skewness and kurtosis in financial time series modeled using SDDEs are a consequence of feedback dynamics in the stochastic delay term.

**see §A.3 for survey of delay differential equations (DDE), stochastic differential equations (SDE), and stochastic delay differential equations (SDDE).

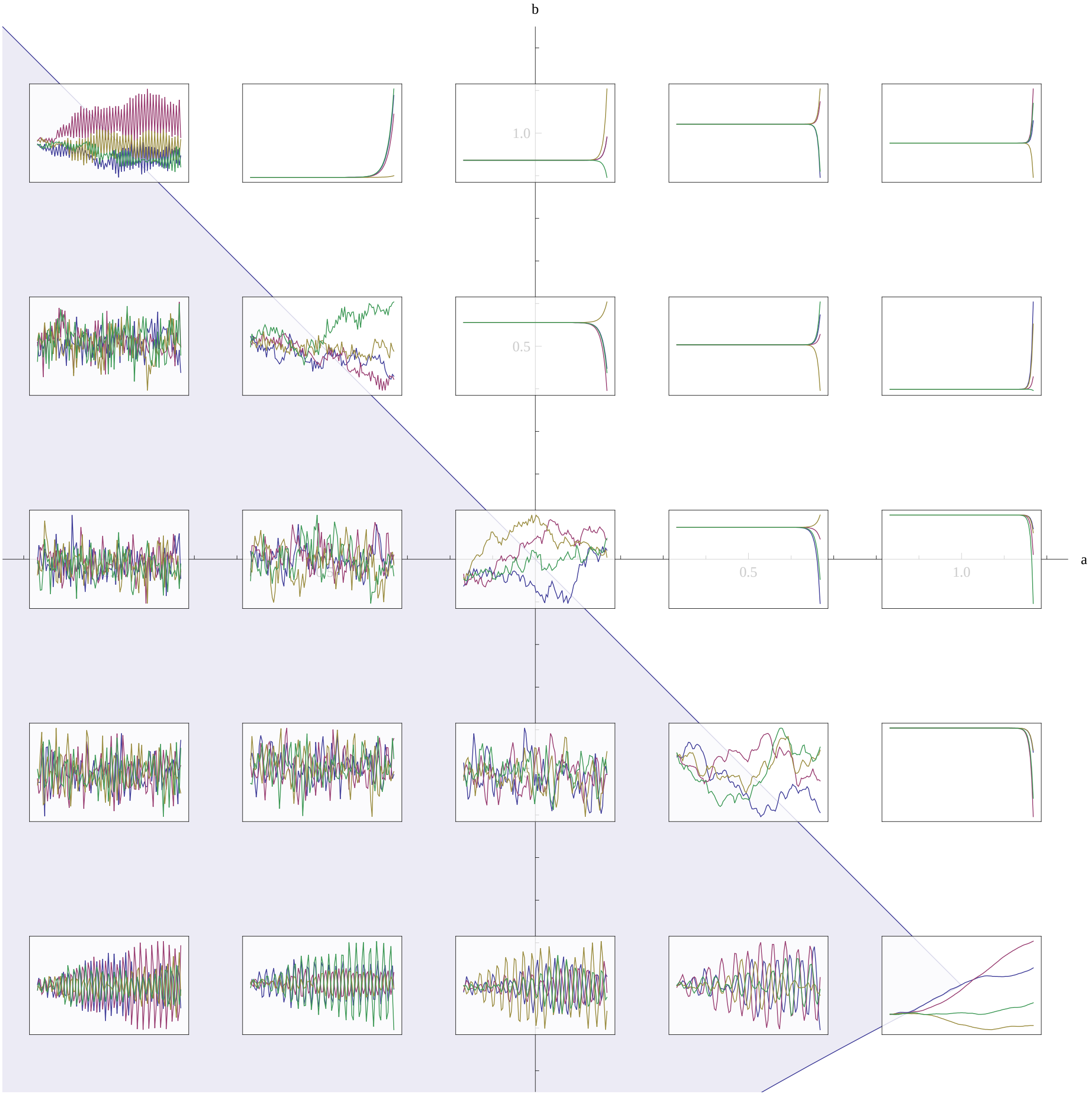


Figure 33: Linear SDDEs $dX(t) = \big(aX(t) + bX(t-r)\big)dt + \sigma dW(t)$

A Appendix

In this Appendix, peripheral topics are included to give the reader a good sense of the background motivating my research work.

First, *related distributions* to the Pearson IV distributions covered so far, ie. normal, t, Nagahara's version of the Pearson IV distribution, and stable distributions are briefly discussed. Further details are available in the original references.

Next, *behavioral finance* explains why and how prices deviate from fundamental values of rational expectations pricing, which can be exploited by profitable option trading strategies. The principal goal of my research was to relax the martingale hypothesis of mathematical finance, ie. that people in financial markets have no memory. For this, I looked at delay differential equations, together with stochastic delay differential equations to account for random effects such as the arrival of news.

Finally, a survey of *delay differential equations (DDE)*, *stochastic differential equations (SDE)* including the Black-Scholes formula, and *stochastic delay differential equations (SDDE)* provides some insight into my motivation to find a multivariate distribution that would be rich enough to capture empirical properties observed in real-world applications.

A.1 Related Distributions

A.1.1 Normal distributions - N($0, 1$), N(μ, σ), and N($\vec{\mu}, \mathbf{\Delta}$)

Definition A.1 (univariate standard and location-scale normal distributions)

Let $x, \mu \in \mathbb{R}$, and $\sigma \in \mathbb{R}^+$. The univariate standard normal distribution has pdf

$$f(x) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}}$$

and the univariate location-scale normal distribution (Figure 34) has pdf

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Definition A.2 (bivariate standard and location-scale normal distributions)

Let $x, y, \mu_1, \mu_2, \rho \in \mathbb{R}$, and $\sigma_1, \sigma_2 \in \mathbb{R}^+$. The bivariate standard normal distribution (Figure 35) has pdf

$$f(x, y) = \frac{1}{2\pi}\ e^{-\frac{x^2+y^2}{2}}$$

and the bivariate location-scale normal distribution (Figure 36) has pdf

$$f(x, y|\mu_1, \mu_1, \sigma_1, \sigma_2, \rho) = \frac{e^{-\frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

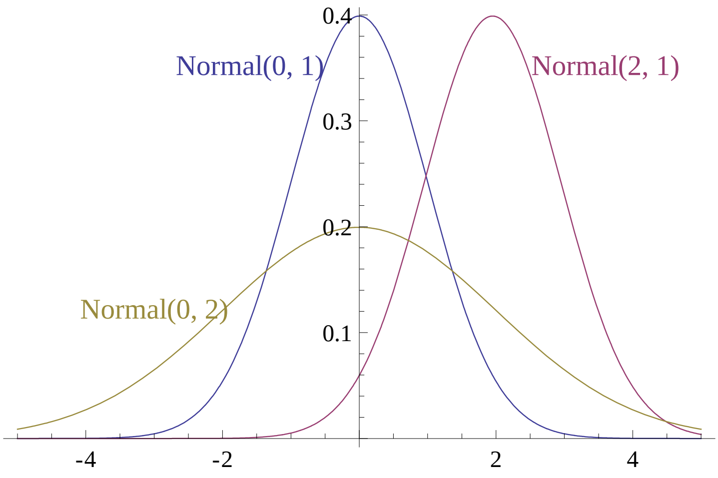


Figure 34: Normal distribution

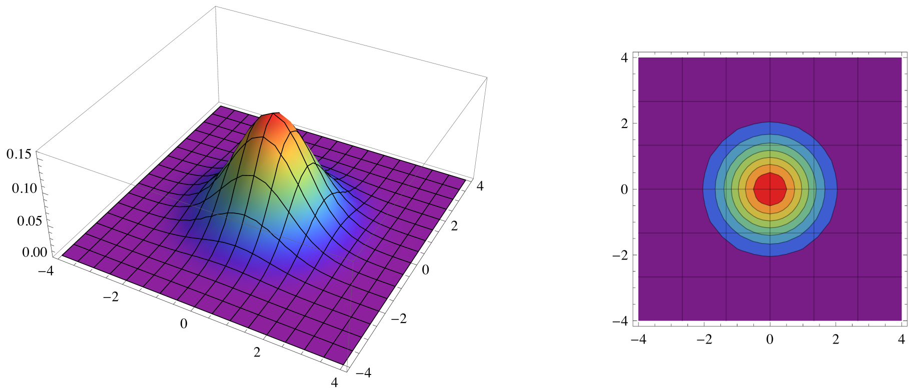


Figure 35: Bivariate normal $N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$

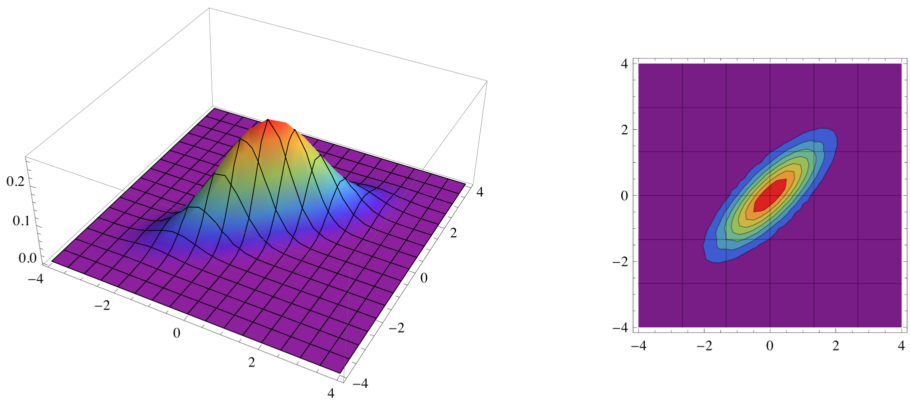


Figure 36: Bivariate normal $N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & .5\\.5 & 1\end{bmatrix}\right)$

Definition A.3 (multivariate normal distribution) *Let $\vec{\boldsymbol{x}}, \vec{\boldsymbol{\mu}} \in \mathbb{R}^N$, and $\boldsymbol{\Delta}$ be a co-variance matrix. Then the multivariate standard normal distribution, eg. trivariate case (Figure 37), has pdf*

$$f(\vec{\boldsymbol{x}} \,|\, \vec{\boldsymbol{0}}, \mathbf{I}) = \frac{1}{(2\pi)^{N/2}}\, e^{-\frac{\vec{\boldsymbol{x}}^T \vec{\boldsymbol{x}}}{2}}$$

and the multivariate location-scale normal distribution has pdf

$$f(\vec{\boldsymbol{x}} \,|\, \vec{\boldsymbol{\mu}}, \boldsymbol{\Delta}) = \frac{1}{(2\pi)^{N/2}\sqrt{|\boldsymbol{\Delta}|}}\, e^{-\frac{(\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})^T \Delta^{-1} (\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})}{2}}$$

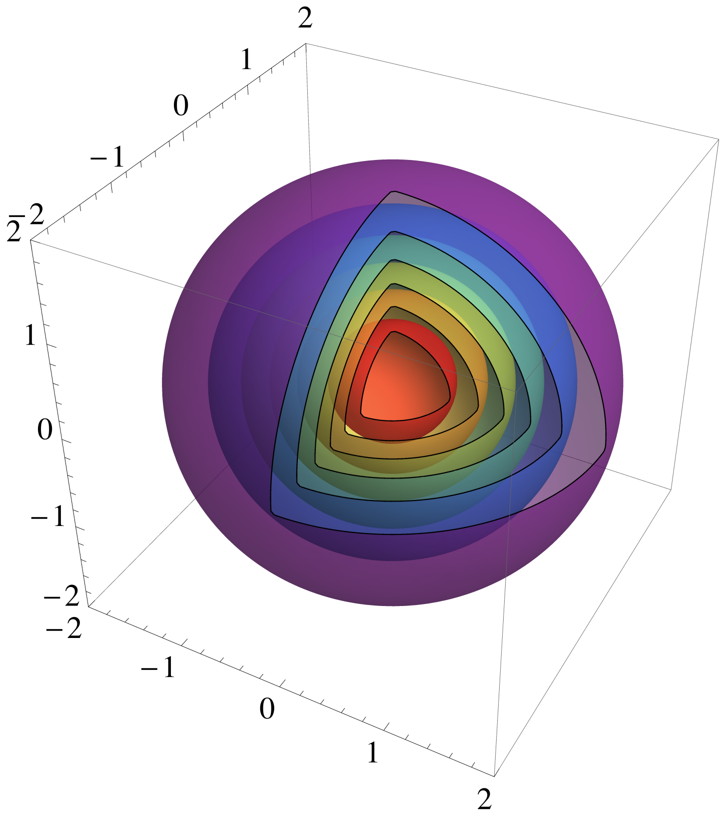


Figure 37: Trivariate normal distribution $\mathrm{N}\left(\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}\right)$

A.1.2 t distributions

A.1.2.1 $t(\nu)$ and $t(\mu, \sigma, \nu)$ - univariate

Definition A.4 (univariate standard and location-scale t distributions)

Let $x, \mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, and $\nu \geq 1$. The univariate standard t distribution has pdf

$$f(x|\nu) = \frac{1}{\sqrt{\nu}\,\mathrm{Beta}\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{43}$$

and the univariate location-scale t distribution has pdf

$$f(x|\mu, \sigma, \nu) = \frac{1}{\sigma\sqrt{\nu}\,\mathrm{Beta}\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{44}$$

Theorem A.5 ($t(\mu, \sigma, \nu)$ moments) *The first four moments of $t(\mu, \sigma, \nu)$ are:*

$$\mu_1 = \begin{cases} \mu, & \nu > 1 \\ \infty, & \nu \leq 1 \end{cases}$$

$$\mu_2 = \begin{cases} \dfrac{\sigma^2 \nu}{\nu - 2}, & \nu > 2 \\ \infty, & \nu \leq 2 \end{cases}$$

$$\mu_3 = \begin{cases} 0, & \nu > 3 \\ \infty, & \nu \leq 3 \end{cases}$$

$$\mu_4 = \begin{cases} \dfrac{3(\nu-2)}{\nu-4}, & \nu > 4 \\ \\ \infty, & \nu \leq 4 \end{cases}$$

Proof. Apply Lemma 2.1 with

$$a = -\mu, \qquad b_0 = \frac{\sigma^2\nu + \mu^2}{\nu+1}, \qquad b_1 = -\frac{2\mu}{\nu+1}, \qquad b_2 = \frac{1}{\nu+1}.$$

□

A.1.2.2 $t_2(\nu)$, $t_2(\rho,\nu)$, and $t_2(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\nu)$ - bivariate

Definition A.6 (bivariate standard, independent, and dependent t distributions)

Let $x, y, \mu_1, \mu_2 \in \mathbb{R}, \sigma_1, \sigma_2 \in \mathbb{R}^+$, and $\nu \geq 1$. The bivariate standard independent t distribution has pdf

$$f(x,y|\nu) = \frac{\Gamma\left(\dfrac{\nu+2}{2}\right)}{\pi\,\nu\,\Gamma\left(\dfrac{\nu}{2}\right)}\left(1+\frac{x^2+y^2}{\nu}\right)^{-\frac{\nu+2}{2}}$$

the bivariate standard dependent t distribution has pdf

$$f(x,y|\rho,\nu) = \frac{\Gamma\left(\dfrac{\nu+2}{2}\right)}{\pi\,\nu\,\Gamma\left(\dfrac{\nu}{2}\right)\sqrt{1-\rho^2}}\left(1+\frac{x^2-2\rho xy+y^2}{\nu}\right)^{-\frac{\nu+2}{2}}$$

and the bivariate location-scale dependent t distribution has pdf

$$f(x, y|\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \nu) =$$

$$\frac{\Gamma\left(\dfrac{\nu+2}{2}\right)}{\pi\,\nu\,\sigma_1\,\sigma_2\,\Gamma\left(\dfrac{\nu}{2}\right)\sqrt{1-\rho^2}} \left(1 + \frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{\nu}\right)^{-\frac{\nu+2}{2}}$$

A.1.2.3 $t_N(\vec{\boldsymbol{\mu}}, \boldsymbol{\Delta}, \nu)$ - multivariate

Definition A.7 (multivariate t distribution)

Let $\vec{\boldsymbol{x}}, \vec{\boldsymbol{\mu}} \in \mathbb{R}^N$, $\boldsymbol{\Delta}$ be a covariance matrix and $\nu \geq 1$. The multivariate location-scale t distribution has pdf

$$f(\vec{\boldsymbol{x}}|\vec{\boldsymbol{\mu}}, \boldsymbol{\Delta}, \nu) = \frac{\Gamma\left(\dfrac{\nu+N}{2}\right)}{(\pi\nu)^{N/2}\Gamma\left(\dfrac{\nu}{2}\right)\sqrt{|\boldsymbol{\Delta}|}} \left(1 + \frac{(\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})^T\boldsymbol{\Delta}^{-1}(\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})}{\nu}\right)^{-\dfrac{\nu+N}{2}}$$

Kotz and Nadarajah (2004) recently wrote a monograph on this multivariate distribution. Genz and Bretz (2009) survey numerical methods for the multivariate normal and multivariate t distributions.

A.1.3 Nagahara's Pearson IV

The Pearson Type IV probability density function is given by

$$f(x)dx = k\left[1+\left(\frac{x-\lambda}{\alpha}\right)^2\right]^{-m}\exp\left[-\nu\tan^{-1}\left(\frac{x-\lambda}{\alpha}\right)\right]dx \qquad (m>1/2) \quad (45)$$

where m, ν, α, and λ are real-valued parameters, and $-\infty < x < \infty$ (k is a normalization constant that depends on m, ν, and α). By Lemma 2.1, the moments can be calculated without knowing k. The mean μ_1 of the pdf is

$$\mu_1 = \lambda - \frac{\alpha\nu}{2(m-1)} \quad (m>1)$$

The variance μ_2 is given by

$$\mu_2 = \frac{\alpha^2}{r^2(r-1)}\left(r^2+\nu^2\right) \qquad (m>3/2)$$

using r for $2(m-1)$. Similarly, the third and fourth moments are

$$\mu_3 = -\frac{4\alpha^3\nu\left(r^2+\nu^2\right)}{r^3(r-1)(r-2)} \qquad (m>2)$$

and

$$\mu_4 = \frac{3\alpha^4\left(r^2+\nu^2\right)\left[(r+6)\left(r^2+\nu^2\right)-8r^2\right]}{r^4(r-1)(r-2)(r-3)} \qquad (m>5/2)$$

The cumulative distribution

$$P(x) = \int_{-\infty}^{x} f(t)dt$$

$P(x)$ can be expressed in terms of the hypergeometric function ${}_2F_1$, which for $|z| \leq 1$ can be calculated via the series

$$
{}_2F_1(\alpha, b; c; z) =
$$

$$
1 + \frac{\alpha b}{c}\frac{z}{1!} + \frac{\alpha(\alpha+1)b(b+1)}{c(c+1)}\frac{z^2}{2!} + \frac{\alpha(\alpha+1)(\alpha+2)b(b+1)(b+2)}{c(c+1)(c+2)\frac{z^3}{3!}} + \cdots
$$

There are many schemes for calculating the hypergeometric function other than direct summation of the power series.

Willink (2008) derives closed-form expressions for the integration constant k and the cumulative distribution function of (45). Using an identity of Stein (1986), Loh (2004) gave an exact expression for the characteristic function of (45) in terms of confluent hypergeometric functions.

A.1.4 Stable Distributions - $Stable(\alpha, \beta, \gamma, \delta)$

Stable distributions (Ibragimov et al., 1971; Zolotarev, 1986; Nolan, 2008)[††] also have non-normal skewness and kurtosis. However, they do not have closed-form expressions for their probability and cumulative distribution functions, making them difficult to use in practice. A stable distribution with parametrization $Stable(\alpha, \beta, \gamma, \delta)$ is defined by its characteristic function

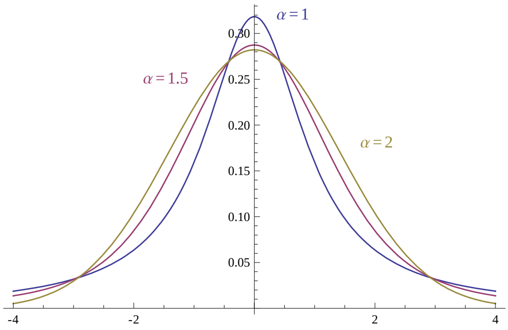
$$\phi(u) = \begin{cases} e^{i\,u\,\delta - \gamma^{\alpha}|u|^{\alpha}\left(1 - i\,\beta\,\frac{u}{|u|}\tan\left(\frac{\pi\alpha}{2}\right)\right)} & \alpha \neq 1 \\ e^{i\,u\,\delta - \gamma|u|\left(1 + i\,\beta\,\frac{u}{|u|}\frac{2}{\pi}\log|u|\right)} & \alpha = 1 \end{cases}$$

for characteristic exponent $\alpha \in (0, 2]$, skewness parameter $\beta \in [-1, 1]$, scale parameter $\gamma > 0$, and location parameter $\delta \in \mathbb{R}$ (Figure 38). Since α and β determine the form of a stable distribution, they are known as *shape parameters*.

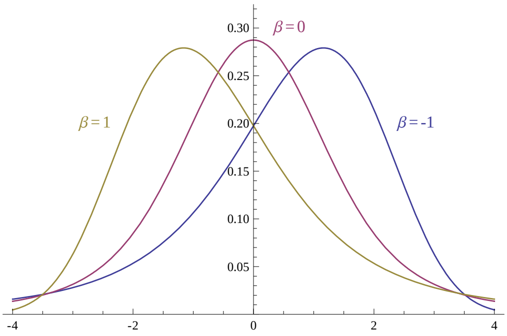
The stable probability distribution function $f(x|\alpha, \beta, \gamma, \delta)$ is obtained via the inverse Fourier transform of its characteristic function

$$f(x|\alpha, \beta, \gamma, \delta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\,u\,x}\phi(u)\,du$$

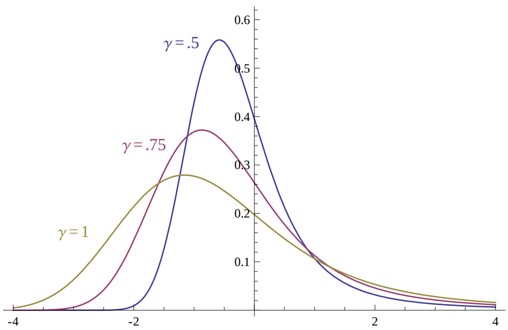
[††]see http://academic2.american.edu/~jpnolan/stable/StableBibliography.pdf for a comprehensive bibliography.



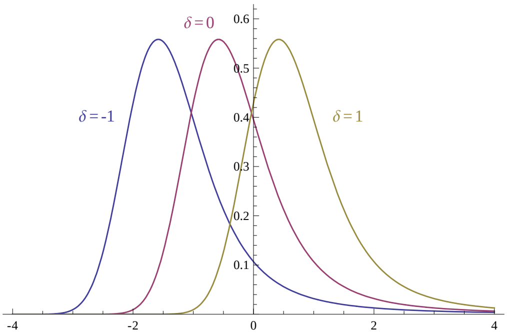
(a) $Stable(\alpha, 0, 1, 0)$



(b) $Stable(1.5, \beta, 1, 0)$



(c) $Stable(1.5, 1, \gamma, 0)$



(d) $Stable(1.5, 1, .5, \delta)$

Figure 38: Stable Distribution

and the stable cumulative distribution function $F(x|\alpha,\beta,\gamma,\delta)$ is therefore

$$F(x|\alpha,\beta,\gamma,\delta) = \frac{1}{2\pi}\int_{-\infty}^{x}\int_{-\infty}^{\infty} e^{-iut}\phi(u)\,du\,dt$$

Note that the Normal and Cauchy distributions are special cases of stable distributions

$$N(\mu,\sigma^2) \;\sim\; Stable(2, 0, \sigma/\sqrt{2}, \mu)$$

$$\text{Cauchy}(\gamma,\delta) \;\sim\; Stable(1, 0, \gamma, \delta).$$

"I've been dealing with these big mathematical models of forecasting the economy, and if I could figure out a way to determine whether or not people are more fearful or changing to euphoric, and have a third way of figuring out which of the two things is working, I don't need any of this other stuff, I could forecast the economy better than anyone I know. Trouble is we can't figure that out. I've been in the forecasting business for 50 years. I'm no better than I ever was and nobody else is. Forecasting 50 years ago was as good or as bad as it is today."

- **Alan Greenspan**, Daily Show with Jon Stewart, 2007/09/18.

A.2 Behavioral Finance

There is an interesting divergence between academic finance and the coverage of financial markets you might see on television. In the real world, financial markets appear to be dominated by people (Figure 39).

Figure 39: NYSE Trading Floor

In contrast, the effects of human behavior seem nearly absent from the academic literature. Why? *Modern finance* is composed of several related theories, including the random walk model of stock prices (Bachelier, 1900), capital asset pricing model (Sharpe, 1964), option pricing (Black and Scholes, 1973), and the fundamental assumption of market efficiency (Samuelson, 1965; Fama, 1970). Investors have no emotions and no memory, they always make *rational* decisions.

Throughout the 1950s and 1960s, economists tried to describe mathematically the emotional state of typical investors (Clark, 2000). They assigned variables to mood, satisfaction, aggressiveness, defensiveness, and safety. These variables were unobservable in the real world. Numbers could not represent vague concepts like expectations of investors, risk aversion, or utility structure. Mathematical models generally can only approximate any application of interest, they rarely can be exact.

Deciding upon which effects are important, and which can be neglected, is as much an art as a science. Difficulties arising from inclusion of various effects can influence such a decision. In an attempt to make finance a "hard" science, the messiness of human behavior was left behind and has been mostly avoided ever since.

Modeling investor behavior is difficult, but not impossible. Human behavior is unlikely to be expressible in closed analytical form. Numerical methods, and inverse problem methods, can be applied to historical data to formulate empirical models that more closely approximate reality than the classical lognormal random walk model.

The traditional finance paradigm seeks to understand financial markets using models in which agents are "rational". First, when they receive new information, agents update their beliefs correctly consistent with Bayes' law of conditional probability. Second, given their beliefs, agents make choices consistent with expected utility.

This simple traditional framework would be very satisfying if its predictions were confirmed in the data. Unfortunately, empirical facts about the aggregate stock market, the cross-section of average returns and individual trading behavior are not captured in this framework.

Behavioral finance has emerged in response to difficulties faced by this traditional paradigm, it argues that some financial phenomena can be better understood using models in which some agents are not fully rational.

The classic objection to behavioral finance argues that even if some agents in the economy are less than fully rational, rational agents will prevent them from influencing security prices for very long, through a process known as *arbitrage*. Strictly speaking, an arbitrage is an investment strategy that offers riskless profits at no cost. In an economy where rational and irrational traders interact, irrationality can have a substantial and long-lived impact on prices. In contrast to textbook arbitrage, real world arbitrage entails both costs and risks, which under some conditions will limit arbitrage and allow deviations from fundamental value to persist. This is known in the literature as "limits to arbitrage".

Behavioral finance rests on the two pillars of *limits to arbitrage* and *investor psychology*, an idea originally due to Shleifer and Summers (1990).

A.2.1 Limits to Arbitrage

While irrational traders are often known as "noise traders" (Black, 1986), rational traders are typically referred to as "arbitrageurs". Noise trader risk, an idea introduced by de Long et al. (1990a) and discussed further in Shleifer and Vishny (1997), is the risk that the mispricing being exploited by the arbitrageur worsens in the short run.

Transaction costs such as commissions, bid-ask spreads, price impact, and the cost of finding and learning about a mispricing, can make it less attractive to exploit such a mispricing (Merton, 1987). Murphy (2006), using a percentage measure of market efficiency, argued that the significant decline in transaction costs provides a plausible explanation for the increase in market efficiency observed over the past decade. This decline in transaction costs is due to the emergence of online trading, and fierce competition seen among several discount brokers. Trading strategies, previously not profitable after transaction costs, can now exploit smaller market inefficiencies.

Arbitrageurs, such as hedge funds, may prefer to trade in the same direction as the noise traders, thereby exacerbating the mispricing, rather than against them. For example, de Long et al. (1990b) *"consider an economy with positive feedback traders, who buy more of an asset this period if it performed well last period. If noise traders push an asset price above fundamental value, arbitrageurs do not sell or short the asset. Rather they buy it, knowing that the earlier price rise will attract more feedback traders next period, leading to still higher prices, at which point the arbitrageurs can exit at a profit."* (Barberis and Thaler, 2002, p.7).

Barberis and Thaler (2002, p.7) continues: *"Hedge funds are not the only market participants trying to take advantage of noise traders. [...] If firm managers believe investors are overvaluing their firm's shares, they can benefit current shareholders by issuing additional shares to the market. This extra supply might push prices back down to fundamental value."* Conversely, if firm managers believe investors are undervaluing their firm's shares, they can repurchase shares from the market, pushing prices back up to fundamental value. This is why markets usually perceive stock buybacks to be a positive signal about a given stock.

Barberis and Thaler (2002, p.8) provides another example of deviation from fundamental value: *"In 1907, Royal Dutch and Shell [...] agreed to merge their interests on a 60:40 basis while remaining separate entities. Shares of Royal Dutch, primarily traded in the United States and in the Netherlands, are a claim to 60 percent of the total cash flow of the two companies. [Shell, primarily traded in the United Kingdom, is a claim to the remaining 40 percent.] If prices equal fundamental value, market* *value of Royal Dutch equity should always be 1.5 times market value of Shell equity."* Relative to efficient benchmark of 1.5, Royal Dutch has traded historically by as much as 35% underpriced relative to parity, and as much as 15% overpriced (Figure 40).

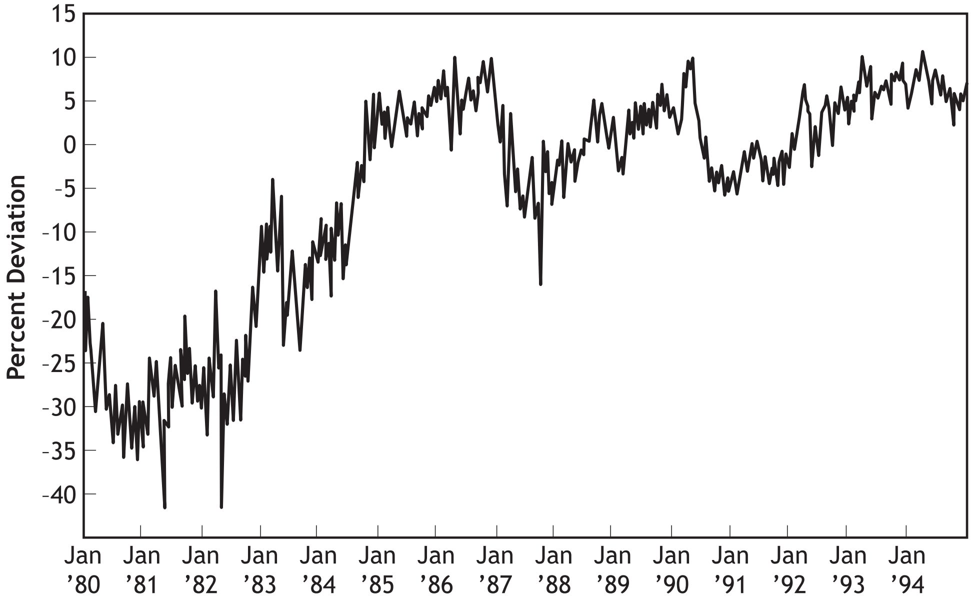


Figure 40: Log deviations from Royal Dutch/Shell parity (Froot and Dabora, 1999).

Since one share is a good substitute for the other, fundamental risk is nicely hedged. Any news about fundamentals affect the two shares equally, leaving the arbitrageur immune. The one risk that remains is noise trader risk. Whatever investor sentiment is causing one share to be undervalued relative to the other could also cause that share to become even more undervalued in the short term.

A.2.2 Investor Psychology

In order to say more about such deviation from fundamental value, behavioral models often assume a specific form of irrationality, taken from extensive experimental evidence compiled by cognitive psychologists on the systematic biases that arise when people form beliefs and preferences (Kahneman et al., 1982; Kahneman and Tversky, 2000; Gilovich et al., 2002).

- *Overconfidence.* Extensive evidence shows that people are overconfident in their judgments. First, confidence intervals people assign to their estimates of quantities are far too narrow. Their 98 percent confidence intervals, for example, include the true quantity only about 60 percent of the time (Alpert and Raiffa, 1982). Second, people are poorly calibrated when estimating probabilities: events they think are certain to occur actually occur only about 80 percent of the time, and events they deem impossible occur about 20 percent of the time (Fischhoff et al., 1977).

  Overconfidence may stem from two other biases, self-attribution bias and hindsight bias. *Self-attribution bias* refers to people's tendency to ascribe any success they have in some activity to their own talents, while blaming failure on bad luck, rather than on their ineptitude. For example, investors might become overconfident after several quarters of investing success (Gervais and Odean, 2001). Hindsight bias is the tendency of people to believe, after an event has occurred, that they predicted it before it happened. Just because someone flips a coin that comes up heads, it does not imply that this person had to ability to make it come up heads.

- *Optimism and Wishful Thinking.* Most people display unrealistically rosy views of their abilities and prospects (Weinstein, 1980). Typically, over 90 percent of those surveyed think they are above average in such domains as driving skill, ability to get along with people, and sense of humor. They also display a systematic planning fallacy: they predict that tasks will be completed much sooner than they actually are (Buehler et al., 1994).

- *Representativeness.* Tversky and Kahneman (1974) show that when people try to determine the probability that data A was generated by model B, or that object A belongs to class B, they often use a representativeness heuristic. They evaluate the probability by how much A reflects the essential characteristics of B.

  Most of the time, representativeness is a helpful heuristic, but it can generate some severe biases. The first is *base rate neglect*. People apply Bayes law

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{46}$$

  incorrectly, putting too much weight on $P(A|B)$, which captures representativeness, and too little weight on the base rate, $P(B)$.

  Representativeness also leads to another bias, *sample-size neglect*. When judging the likelihood that a data set was generated by a particular model, people often fail to take the size of the sample into account. A small sample is seen to be as representative as a large one. In cases where people do not initially know the data-generating process, they tend to infer it too quickly on the basis of too few data points. This belief that even small samples will reflect the properties of the parent population is sometimes known as the *law of small numbers* (Rabin, 2002). In situations where people *do* know the data-generating process in advance, the law of small numbers leads to a gambler's fallacy effect. If a fair coin generates five heads in a row, people will say that "tails are due". They believe that even a short sample should be representative of the fair coin, so there has to be more tails to balance out the large number of heads.

- *Belief Perseverance.* Once people have formed an opinion, they cling to it too tightly and for too long (Lord et al., 1979). First, people are reluctant to search for evidence that contradicts their beliefs. Second, even if they find such evidence, they treat it with excessive skepticism. In the context of academic finance, belief perseverance explains why people, who start out believing in the Efficient Markets Hypothesis, continue to believe in it long after compelling evidence to the contrary has emerged.

Further biases, such as *conservatism*, *anchoring*, and *availability*, are discussed in Barberis and Thaler (2002).

An essential ingredient of any model trying to understand asset prices or trading behavior is an assumption about investor preferences, or about how investors evaluate risky gambles. Many models assume that investors evaluate gambles according to expected utility theory (von Neumann and Morgenstern, 1953). However, people systematically violate expected utility theory when choosing among risky gambles.

Prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) may be the most promising for financial applications (Benartzi and Thaler, 1995; Barberis et al., 2001). Many other theories have been proposed, such as *weighted-utility theory* (Chew and MacCrimmon, 1979; Hong, 1983), *implicit expected utility* (Dekel, 1986; Chew, 1989), *disappointment aversion* (Gul, 1991), *regret theory* (Bell, 1982; Loomes and Sugden, 1982), *rank-dependent utility theories* (Quiggin, 1982; Segal, 1987, 1989; Yaari, 1987). However, prospect theory is the most successful at capturing experimental results.

Kahneman and Tversky (1979) layed out the original version of prospect theory, designed for gambles with at most two nonzero outcomes. When offered a gamble

$$(\text{outcome } x, \text{probability } p; \quad \text{outcome } y, \text{probability } q),$$

people assign it a value of

$$\pi(p)v(x) + \pi(q)v(y),$$

where v and π are shown in Figure 41. When choosing between different gambles, they pick the one with the highest value.

An important feature is the shape of the value function v, namely its concavity in the domain of gains and convexity in the domain of losses. People are risk-averse over

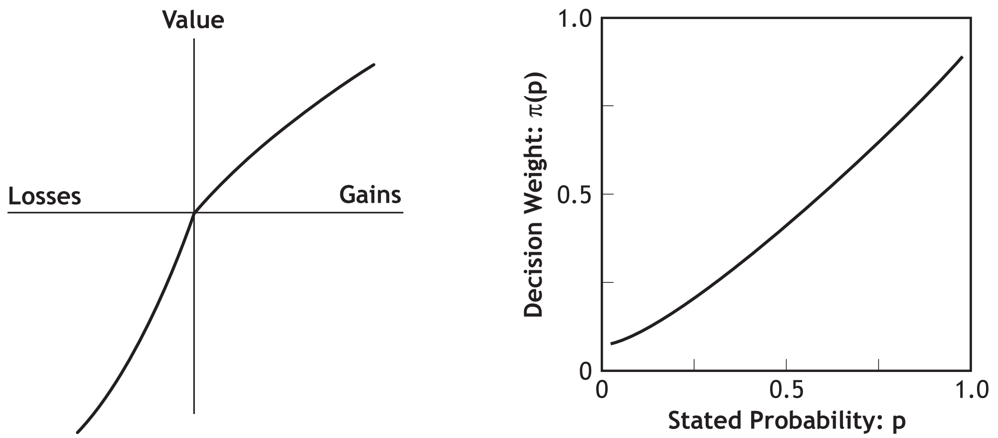


Figure 41: Kahneman and Tversky (1979) value and probability weighting functions.

gains, and risk-seeking over losses, eg.

$$\left(1000, \frac{1}{2}\right) \prec \left(500, 1\right)$$

but

$$\left(-1000, \frac{1}{2}\right) \succ \left(-500, 1\right)$$

where $A \succ B$ denotes "*A is more popular than B*", formally defined as "*a statistically significant fraction of subjects preferred A to B*". In other words, people settled for a certain small gain over a potential large gain, but risked a potential large loss to avoid a certain small loss.

Tversky and Kahneman (1992) extend prospect theory to gambles with more than two outcomes. If a gamble promises outcome x_i with probability p_i, they propose that people assign this gamble the value

$$\sum_i \pi_i \, v(x_i)$$

where

$$v(x_i) = \begin{cases} x_i^\alpha, & x_i \geq 0 \\ -\lambda(-x_i)^\alpha, & x_i < 0 \end{cases}$$

and

$$\pi_i = \frac{p_i^\gamma}{\left(p_i^\gamma + (1-p_i)^\gamma\right)^{1/\gamma}}$$

Based on experimental evidence, Tversky and Kahneman (1992) estimate

$$\hat{\alpha} = 0.88, \qquad \hat{\lambda} = 2.25, \qquad \hat{\gamma} = 0.65$$

where λ is the coefficient of loss aversion, ie. a measure of the relative sensitivity to gains and losses.

Framing refers to the way a problem is posed to a decision maker, who also has flexibility in how to think about the problem. For example, suppose a gambler wins \$200 on his first bet, but loses \$50 on his second bet. Does he value the outcome of the second bet as a loss of \$50, or as a reduction to the previous gain of \$200? The process by which people formulate such problems for themselves is called *mental accounting* (Thaler, 1999), which matters in prospect theory since v is nonlinear.

People dislike situations where they are uncertain about the probability distribution of a gamble (Ellsberg, 1961), something known as *ambiguity aversion*. An early discussion of this aversion can be found in Knight (1921), who defined *risk* as a gamble with known distribution and *uncertainty* as a gamble with unknown distribution. He suggested that people dislike uncertainty more than risk.

Stocks have historically earned much higher risk-adjusted returns than bonds, something known as the *equity premium puzzle* (Hansen and Singleton, 1983; Mehra and Prescott, 1985). Benartzi and Thaler (1995) link prospect theory to the equity premium. They study how an investor with prospect theory preferences allocates his financial wealth between stocks and bonds. If investors get utility from annual changes in financial wealth and are loss averse over these changes, their fear of a major drop in financial wealth will lead them to demand a higher premium as compensation. Equity premium might be overstated due to survivorship bias (Brown et al., 1995).

The historical volatility of stock returns, known as the *volatility puzzle* (Campbell, 1999), is difficult to explain with any model in which investors are rational and discount rates are constant (Shiller, 1981; LeRoy and Porter, 1981).

A.2.3 Momentum

Jegadeesh and Titman (1993) show that stocks that perform the best (worst) over a three- to twelve-month period tend to continue to perform well (poorly) over the subsequent three to twelve months.

Grinblatt et al. (1995) and Chen et al. (2000) find that mutual funds tend to buy past winners and sell past losers. Also, Womack (1996) and Jegadeesh et al. (2004) report that analysts generally recommend high-momentum stocks more favorably than low momentum stocks.

Griffin et al. (2003) find that the momentum strategies are profitable in North America, Europe, and Latin America, but they are not significantly profitable in Asia. Are Asians investors less emotional about their stocks?

Momentum profits can also potentially arise if stock prices react to common factors with some delay. Intuitively, if stock prices react with a delay to common information, investors will be able to anticipate future price movements based on current factor realizations and devise profitable trading strategies. In some situations such delayed reactions will result in profitable contrarian strategies and in some other situations, it will result in profitable momentum strategies. To see this, consider the return generating process

$$r_{it} = \mu_i + \beta_i f_t + \gamma_i f_{t-1} + \epsilon_{it}$$

where β_i and γ_i are sensitivities to contemporaneous and lagged factor realizations.

Many authors, including Lo and MacKinlay (1990); Brennan et al. (1993); Jegadeesh and Titman (1993, 1995) have used this delayed reaction model to characterize stock return dynamics.

A.2.4 Farmer's (2002) Market

Farmer (2002) studies the long-term evolution of markets as a dynamical system of money flows. Capital is allocated to various investment strategies as a function of previous performance. Market inefficiencies are eliminated as the strategies that exploit them accumulate capital. Strategies grow by reinvesting profits, and/or attracting additional capital based on track record. Eventually, market impact of these strategies cancel out the original pattern, as demonstrated by the decline over recent years of the well-known *January Effect* (Gu, 2003). Given that lengthy track records are needed to attract significant investment capital, progression to market efficiency can take months, or even years.

Farmer (2002) proposes the price formation process

$$\frac{dP}{dt} = f\big(D(P)\big)$$

where P is the price at time t, D is the excess demand (demand-supply), and f is an increasing function known as the *market impact function*.

The rate of change of the price increases as there is excess demand for a security. Conversely, the rate of change of the price decreases when there is excess supply.

Excess demand is a function of price so that different prices will change demand and supply for a security. Feedback between P and its derivative $\frac{dP}{dt}$ leads to complex price dynamics.

Decision strategies, employed by investors and traders, can be classified into three broad groups. *Technical* or *chartist* strategies, of which *trend-following* is a special case, depend only on past prices. *Value* or *fundamental* strategies are based on a subjective assessment of value in relation to price.

Value investors buy when they believe an asset is undervalued and sell, or take a short position, when they see an asset as overvalued. As a result, value strategies induce negative short-term autocorrelations in price. Trend-following strategies, conversely, induce positive short-term autocorrelations since trend-following investors buy when the price of an asset has been recently rising and sell when the price of an asset has been declining.

Let there be N traders holding $x_i(t)$ shares at time t, and let the position of trader i be a function

$$x_i(t+1) = S_i\big(P(t), P(t-1), \ldots, I_i(t)\big)$$

where $I_i(t)$ represents any additional external information. The function S_i is the strategy, or decision rule, of agent i.

Without loss of generality, we can assume strategies of all traders are distinct. If k agents $i_1, i_2, \ldots, i_k$ are all using the same strategy, then they are equivalent to a single "representative agent" with position $\sum_{j=1}^{k} x_{i_j}(t)$. The buy/sell order $\omega_i(t)$ of agent i is the derivative of its position

$$\omega_i(t) = x_i(t) - x_i(t-1) \tag{47}$$

Excess demand $D\big(P(t)\big)$ from (47) can be represented as the net order

$$\omega = \sum_{i=1}^{N} \omega_i$$

An approximation of the market impact function can be derived by assuming that it is of the form

$$P(t+1) = f\big(P(t), \omega\big) = P(t)\phi(\omega)$$

where ϕ is an increasing function with $\phi(0) = 1$. Taking logarithms and expanding in a Taylor's series, provided derivative $\phi'(0)$ exists, to leading order

$$\log P(t+1) - \log P(t) \approx \frac{\omega}{\lambda} \tag{48}$$

This functional form for ϕ is called the *log-linear market impact function*. λ is a scale factor that normalizes the order size, and is called the *liquidity*.

We can now describe the interaction between trading decisions and prices as a dynamical system. Let $p(t) = \log P(t)$, and adding a random term $\xi(t)$ for external perturbations not driven by trading such as news announcements, (48) becomes

$$p(t+1) = p(t) + \frac{1}{\lambda}\sum_{i=1}^{N}\omega_i\Big(P(t), P(t-1), \ldots, I_i(t)\Big) + \xi(t+1) \tag{49}$$

"Each trading strategy influences prices [on small timescales], and in turn prices influence each trading strategy [on larger timescales]" (Farmer, 2002, p.907). The profit $g_i(t)$ of agent i at time t is

$$g_i(t) = \Big(\Delta P(t) + d(t)\Big)x_i(t-1) \tag{50}$$

where $d(t)$ accounts for a dividend payment at time t. Substitute for $\Delta P(t)$ from (49) and $\omega(t)$ from (47) to get

$$g_i(t) = \left(\frac{1}{\lambda}\sum_{i=1}^{N}\big(x_i(t) - x_i(t-1)\big) + \xi(t) + d(t)\right)x_i(t-1) \tag{51}$$

Farmer (2002, p.930) continues: *"Decisions about capital allocation are made by human beings, making them difficult to model exactly. However, the human [...] tendency to rely on strategies that have been successful in the past"* can be described by a dynamical system equivalent to the standard Lotka-Volterra model of population biology.

Some important factors that influence capital allocations are *reinvestment of earnings* (a fraction of profits is added to existing capital), *attracting capital from investors* (investors tend to allocate funds to money managers who have been successful in the past), and *other economic necessity* (investors might withdraw funds for reasons unrelated to past performance, eg. for personal consumption). A simple model that captures all three factors is

$$\Delta c_i(t) = a_i g_i(t-1) + \gamma_i \tag{52}$$

where $c_i(t)$ is the capital for agent i at time t, and $g_i(t)$ is the profit. Farmer (2002, p.931) explains (52): *"The profit is lagged by one because it is not available for reinvestment until the following period. a_i is the reinvestment rate. Because of the ability to attract outside capital, which depends on past performance, it is possible that $a_i > 1$"*, ie. a_i captures the first two factors, while γ_i captures the third.

Substitute (51) into (52) to get

$$\Delta c_j(t{+}1) = \sum_{i=1}^{N}\Big(A_{ij}(t)c_i(t)c_j(t{-}1) - B_{ij}(t)c_i(t{-}1)c_j(t{-}1)\Big) + \mu_j(t)c_j(t{-}1) + \gamma_j \tag{53}$$

where

$$A_{ij}(t) = \frac{a_j}{\lambda}x_i(t)x_j(t{-}1), \quad B_{ij}(t) = \frac{a_j}{\lambda}x_i(t{-}1)x_j(t{-}1), \quad \mu_j(t) = \big(\xi(t)+d(t)\big)x_j(t{-}1)$$

(53) is a *"quadratic difference equation with time-varying coefficients"* (Farmer, 2002, p.931). If reinvestment is sufficiently slow, then (53) can be rewritten as a

differential equation of the form

$$\dot{c}_j(t) = \frac{a_j}{\lambda} \sum_{i=1}^{N} G_{ij}\, c_i(t) c_j(t) + \mu_j(t) c_j(t) + \gamma_i \tag{54}$$

The *gain matrix* G describes the profits due to interactions with other strategies, and μ_j describes those due to correlations with external fluctuations and dividend payments.

- *Competition ($G_{ij} < 0$ and $G_{ji} < 0$).* *"Crowding lowers individual profits. [...] The capital associated with the maximum profit level can be thought of as the carrying capacity of the strategy"* (Farmer, 2002, p.929).

- *Predator-Prey ($G_{ij} > 0$ and $G_{ji} < 0$).* Strategy i is taking money from strategy j, ie. it "preys" on j. For example, momentum traders buying early in a cycle will profit by selling their shares to other momentum traders entering late in the cycle.

- *Symbiosis ($G_{ij} > 0$ and $G_{ji} > 0$).* Some strategies mutually benefit from the presence of others. In the above example, early momentum traders reinforce an upward movement in prices to overshoot fundamental values, creating an opportunity for contrarian value investors. The same momentum traders then benefit from the new downward momentum cycle created by contrarian value investors becoming active in the market.

(54) are the generalized Lotka-Volterra equations (Murray, 1990; Hofbauer and Sigmund, 1998), originally introduced to model dynamics in populations with predator-prey relationships. For financial markets, population is replaced by capital. Dynamics of Lotka-Volterra equations can be stable or unstable, with fixed points, limit cycles or chaotic attractors. Fixed points are obtained by setting the left-hand side of (54) to zero, which gives a system of coupled quadratic equations for c_i with N unknowns. Such equations can have as many as 2^N roots and the possibility of multiple equilibria.

In summary, the Farmer's market can be written as

$$\left\{ \begin{array}{c} p(t+1) = p(t) + \dfrac{1}{\lambda}\sum_{i=1}^{N}\Big(c_i(t+1)x_i(t+1) - c_i(t)x_i(t)\Big) + \xi(t+1) \\ \\ c_i(t+1) = c_i(t) + a_i g_i(t) + \gamma_i \\ \\ g_i(t) = \Big(P(t) - P(t-1) + d(t)\Big)c_i(t-1)x_i(t-1) \end{array} \right\}$$

where

- $P(t)$ price, $p(t) = \log P(t)$;
- $x_i(t)$ position of agent i;
- $c_i(t)$ capital of agent i;
- $g_i(t)$ profits reinvested at rate a_i;
- $d(t)$ dividend paid at time t;
- γ_i net capital inflow of agent i.

A.3 Differential Equations

A.3.1 Delay Differential Equations

Ordinary differential equations (ODE) are commonly used to model natural phenomena. Future behavior is uniquely determined by the present state of a system and independent of the past. In many applications, past state(s) also influence the future of this system in a significant manner. Neglecting such delays can lead to false conclusions, eg. presence of small delays altering the stability properties of a given system (Kolmanovskii and Myshkis, 1999, p. 351). At least, we are missing an opportunity to obtain a model that more closely approximates the application of interest.

Example A.8 (DDE - population growth model with delay) *Consider a population composed of adult and newborn individuals. Let $N(t)$ denote the density of adults at time t, and let h be the maturation period of a newborn to adulthood. Adults produce offspring at a per-capita rate of α and die at a rate of μ. Assume that a newborn survives to become an adult with probability ρ, and let $r = \alpha\rho$. Then the dynamics of N can be described by the differential equation*

$$\dot{N}(t) = -\mu N(t) + rN(t-h)$$

which involves a nonlocal term, $rN(t-h)$ meaning that newborns become adults with some delay. In this setting, the time evolution of the population N involves the current as well as the past values of N.

These equations are known as *delay differential equations* (DDE). They are also known as *functional differential equations* (FDE). DDEs have applications in viscoelasticity, mechanics, nuclear reactors, distributed networks, heat flow, neural networks, combustion, interaction of species, microbiology, learning models, epidemiology, physiology, economics, and many others (Kolmanovskii and Myshkis, 1999).

Results concerning existence, uniqueness, continuous dependence on parameters, and so on, are essentially the same as for ODEs. However, many of these proofs face additional technicalities due to an infinite-dimensional setting (Hale et al., 1993; Magalhäaes et al., 2002; Arino et al., 2006). Whereas most of the analysis of ODEs occurs in Euclidean spaces, analysis of DDEs must take place in infinite-dimensional functional spaces (Aliprantis and Border, 1999; Da Prato, 2006). Fortunately, as a consequence of the *method of steps* below, some other proofs and most importantly numerical methods (Shampine and Thompson, 2000, 2001; Bellen and Zennaro, 2003; Zivaripiran, 2005) are easily adapted from ODEs to DDEs.

Let's start with the simplest possible DDEs, namely linear equations with constant coefficients.

Definition A.9 (DDE - linear rfde) *linear retarded functional differential equation (RFDE) with constant coefficients*

$$\begin{aligned} \dot{x}(t) &= ax(t) + bx(t-r) + f(t), \qquad t \geq 0 \\ x(t) &= \phi(t), \qquad t \in [-r, 0] \end{aligned} \tag{55}$$

Lemma A.10 (DDE - variation-of-constants formula) *Linear nonhomogeneous scalar ODE*

$$\dot{x}(t) = ax(t) + f(t)$$

has solution

$$x(t) = ce^{at} + \int_0^t e^{a(t-s)} f(s)ds \tag{56}$$

Theorem A.11 (DDE - method of steps) *If ϕ is a continuous function on $[-r, 0]$, then there exists a unique function $x(\phi, f)$ on $[-r, \infty)$ that satisfies (55).*

Proof. 1. On $[0, r]$ "first step",

$$x(t) = \phi(t) \text{ on } [-r, 0] \tag{57}$$

$$\Rightarrow \quad x(t-r) = \phi(t-r) \text{ on } [0, r] \tag{58}$$

so (55) becomes

$$\left\{ \begin{aligned} & x(t) = ax(t) + \underbrace{b\phi(t-r) + f(t)}_{g(t)}, \qquad t \in [0, r] \\ & x(0) = \phi(0) \end{aligned} \right\} \tag{59}$$

which is a linear nonhomogeneous ODE initial value problem on $[0, r]$. By variation-of-constants formula (56),

$$x(t) = \phi(0)e^{at} + \int_0^t e^{a(t-s)} g(s)ds \text{ on } [0, r]$$

2. Now for $[r, 2r]$ "second step",

$$x(t) \text{ is known on } [0, r]$$

$$\Rightarrow \quad x(t-r) \text{ is known on } [r, 2r]$$

and we get another ODE initial value problem

$$x(t) = ax(t) + \underbrace{bx(t-r) + f(t)}_{h(t)}, \qquad t \in [r, 2r]$$

$$x(r) = \underbrace{\phi(0)e^{ar} + \int_0^r e^{a(r-s)} g(s)ds}_{\text{known constant}}$$

3. Repeat this method of steps to get solution of (55) on $[-r, \infty)$.

□

Theorem A.12 (DDE - characteristic equation) *The linear homogeneous RFDE*

$$x(t) = ax(t) + bx(t-r) \tag{60}$$

has solution

$$x(t) = ce^{\lambda t} \tag{61}$$

if and only if

$$h(\lambda) := \lambda - a - be^{\lambda r} = 0$$

where $h(\lambda)$ is known as the characteristic equation of (60).

Proof.

$$(60) \quad \Rightarrow \quad \dot{x}(t) = ace^{\lambda t} + bce^{\lambda(t-r)}$$

$$(61) \quad \Rightarrow \quad \dot{x}(t) = \lambda ce^{\lambda t}$$

$$\therefore \quad \lambda - a - be^{\lambda r} = 0$$

□

Theorem A.13 (DDE - fundamental solution) *Let $X(t)$ satisfy*

$$\left\{ \begin{array}{l} \dot{X}(t) = aX(t) + bX(t-r), \qquad t \geq 0 \\ \\ X(t) = 0, \qquad t < 0 \\ \\ X(0) = 1 \end{array} \right\} \tag{62}$$

Then $X(t)$ is the fundamental solution, ie.

$$\mathcal{L}(X)(\lambda) = \frac{1}{h(\lambda)}$$

where $\mathcal{L}(f)(\lambda)$ is the well-known Laplace transform defined by

$$\mathcal{L}(f)(\lambda) = \int_0^\infty e^{-\lambda t} f(t) dt$$

Proof. Multiply (62) by $e^{-\lambda t}$ to get

$$e^{-\lambda t}\dot{X}(t) = ae^{-\lambda t}X(t) + be^{-\lambda t}X(t-r)$$

$$\Rightarrow \quad \underbrace{\int_0^\infty e^{-\lambda t}\dot{X}(t)dt}_{\dagger} = \underbrace{a\int_0^\infty e^{-\lambda t}X(t)dt}_{\mathcal{L}(X)(\lambda)} + \underbrace{b\int_0^\infty e^{-\lambda t}X(t-r)dt}_{\ddagger}$$

$$\dagger \;=\; \left[e^{-\lambda t}X(t)\right]_0^\infty + \lambda\int_0^\infty e^{-\lambda t}X(t)dt \;=\; 0 - \underbrace{X(0)}_{=1 \text{ by } (62)} + \lambda\mathcal{L}(X)(\lambda)$$

$$\ddagger \;=\; b\int_{-r}^\infty e^{-\lambda(r+s)}X(s)ds$$

$$=\; be^{-\lambda r}\left(\int_{-r}^0 e^{-\lambda s}\underbrace{X(s)}_{=0 \text{ by } (62)}ds + \int_0^\infty e^{-\lambda s}X(s)ds\right)$$

$$=\; be^{-\lambda r}\mathcal{L}(X)(\lambda)$$

$$\therefore \quad \mathcal{L}(X)(\lambda)\big(\lambda - a - be^{-\lambda r}\big) = 1$$

$$\Rightarrow \quad \mathcal{L}(X)(\lambda) = \frac{1}{h(\lambda)}$$

Note: Hale et al. (1993, p. 20) uses the confusing notation $h^{-1}(\lambda)$ instead of $\dfrac{1}{h(\lambda)}$.

$\square$

In general, the right-hand derivative $x'(0)^+$ does not equal the left-hand derivative $\phi'(0)^-$. Hence, the solution $x(t)$ is not smoothly joined to the initial function $\phi(t)$ at 0, where only C^0 continuity can be assured. Moreover, such a discontinuity in the derivative propagates periodically to $r, 2r, \ldots$, where the solution is smoothed out to $C^1, C^2, \ldots$

Example A.14 (DDE - rfde smoothing property)

$$\begin{aligned} x'(t) &= -x(t-1), \qquad t \geq 0 \\ x(t) &= 1, \qquad t \in [-1, 0] \end{aligned} \tag{63}$$

whose solution is depicted in Figure 42.

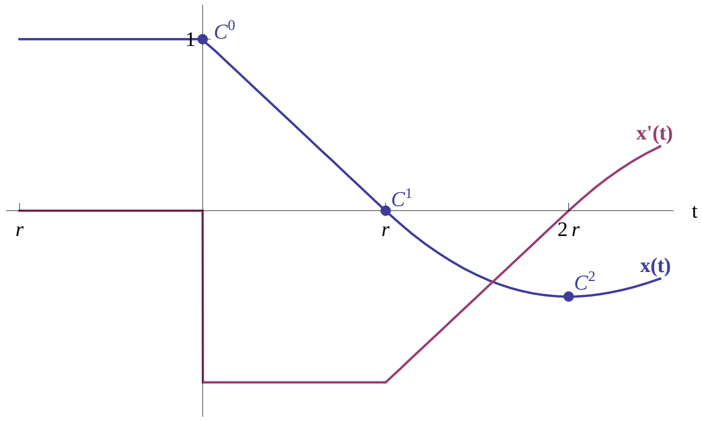


Figure 42: Smoothing Property of RFDE

Definition A.15 (DDE - general rfde) *Let $\sigma \in \mathbb{R}, A \geq 0, x \in C\big([\sigma - r, \sigma + A], \mathbb{R}^n\big)$, the Banach space with sup-norm. For $t \in [\sigma, \sigma + A]$, let $x_t \in C$ defined by*

$$x_t(\theta) = x(t + \theta), \qquad -r \leq \theta \leq 0 \tag{64}$$

Let $D \subseteq \mathbb{R} \times C$, $f : D \to \mathbb{R}^n$, then RFDE is

$$\begin{aligned} x(t) &= f(t, x_t), \qquad t \geq 0 \\ x(t) &= \phi(t), \qquad t \in [-r, 0] \end{aligned} \tag{65}$$

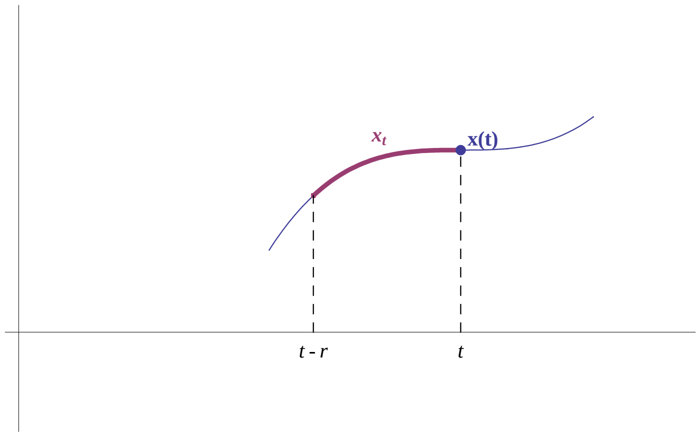


Figure 43: x_t vs. $x(t)$

A.3.2 Stochastic Differential Equations

In addition to deterministic delays discussed in the previous section, one might wish to model random effects disturbing the system by allowing for randomness in some or all of the coefficients of a differential equation[‡‡].

For example, consider the deterministic population growth model

$$\frac{dN}{dt} = a(t)N(t), \qquad N(0) = N_0 \text{ (constant)} \tag{66}$$

where $N(t)$ is the population at time t, and $a(t)$ is the rate of growth at time t. Suppose $a(t)$ is not completely known, but equal to $r(t) + \xi(t)$ where $r(t)$ is known and deterministic. Exact behavior of random noise $\xi(t)$ is not known, except for its probability distribution. The original deterministic differential equation (66) then becomes a *stochastic differential equation* (SDE)

$$\frac{dN}{dt} = \Big(r(t) + \xi(t)\Big)N(t), \qquad N(0) = N_0 \text{ (constant)}$$

which can also be written in differential form

$$dN(t) = r(t)N(t)dt + \alpha N(t)dW(t) \tag{67}$$

where $W(t)$ is the standard Wiener process, and $\xi(t) = \alpha W(t)$.

[‡‡]We will look at the combination of delay and randomness in the next subsection.

What is the solution to (67)? Is it a real-valued function? No, the solution to an SDE is a *stochastic process*. (67) is an example of the general form (68) of an SDE.

Oksendal (2003) is a well-known introductory text to stochastic differential equations. Evans (2006) provides an excellent set of lecture notes on this topic. The standard results presented in this section, and their proofs, are available in many reference texts (Friedman, 1975, 1976; Da Prato and Zabczyk, 1992; Kloeden and Platen, 1992; Da Prato and Tubaro, 2002; Protter, 2003; Cherney and Engelbert, 2005).

Before continuing, we need to define *probability space*, *random variable*, and *stochastic process*.

Definition A.16 (probability space) *A triple* $(\Omega, \mathcal{F}, P)$ *is called a* probability space *given a set* Ω, *a* σ*-field* $\mathcal{F}$*:*

1. $\emptyset, \Omega \in \mathcal{F}$.
2. $A \in \mathcal{F} \quad \Rightarrow \quad A^c \in \mathcal{F}$.
3. $A_1, A_2, \ldots \in \mathcal{F} \quad \Rightarrow \quad \bigcup_{k=1}^{\infty} A_k, \bigcap_{k=1}^{\infty} A_k \in \mathcal{F}$.

and a probability measure $P : \mathcal{F} \to [0, 1]$*:*

1. $P(\emptyset) = 0,\ P(\Omega) = 1$.
2. $A_1, A_2, \ldots \in \mathcal{F} \quad \Rightarrow \quad P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k)$
3. $A_1, A_2, \ldots \in \mathcal{F}$ *and disjoint* $\quad \Rightarrow \quad P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$

A random variable is a function whose output is random and to which a probability distribution is assigned. Formally, a random variable is a measurable function from a sample space to the measurable space of possible values of the variable.

Definition A.17 (random variable) *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A mapping*

$$\mathbf{X} : \Omega \to \mathbb{R}^n$$

is called a random variable if for each $B \in \mathcal{B}$, we have

$$\mathbf{X}^{-1}(B) \in \mathcal{F}$$

where $\mathcal{B}$, known as the Borel σ-field, denotes the smallest σ-field containing all open subsets of $\mathbb{R}^n$.

Definition A.18 (stochastic process)

1. *A collection $\{\mathbf{X}(t) \mid t \geq 0\}$ of random variables is called a stochastic process.*
2. *For each point $\omega \in \Omega$, the mapping $t \mapsto \mathbf{X}(t, \omega)$ is the corresponding sample path.*

Definition A.19 (brownian motion / wiener process) *A real-valued stochastic process $\{W(t) \mid t \geq 0\}$ is called Brownian motion, or Wiener process, if*

1. $W(0) = 0$ *a.s.,*
2. $W(t) - W(s) \sim N(0, t-s)$ *for all* $t \geq s \geq 0$*,*
3. *the increments* $W(t_1), W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1})$ *are independent random variables for all times* $0 < t_1 < t_2 < \cdots < t_n$*.*

The formal time-derivative

$$\dot{W}(t) = \frac{dW(t)}{dt} = \xi(t)$$

is known as white noise.

Definition A.20 (stochastic differential equation (SDE))

$$\left\{ \begin{array}{l} dX(t) = f\big(t, X(t)\big)\ dt + g\big(t, X(t)\big)\ dW(t) \\ \\ X(0) = x_0 \end{array} \right\} \tag{68}$$

We say that a stochastic process $X(t)$ solves (68) provided

$$X(t) = x_0 + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dW(s), \qquad \text{for all } t > 0$$

but first we must define the *stochastic integral* $\int_0^t \cdot\, dW$.

Itô's lemma, a stochastic version of the chain rule, is the key to solving many SDEs. Itô's (1944; 1946) theory of stochastic integration was originally motivated to construct diffusion processes (a subclass of Markov processes) as solutions of stochastic differential equations. Itô's stochastic integral, and the proof to Itô's lemma, is constructed in a similar fashion to the Lebesque integral by starting with simple *step processes*, and then approximating stochastic processes as limits of step processes. *Stochastic calculus* can be considered as the theory of differentiation and integration of stochastic processes, as opposed to functions in deterministic calculus.

Theorem A.21 (SDE - Itô's lemma) *Let $X(t)$ be a real-valued stochastic process with stochastic differential*

$$dX(t) = F(t)dt + G(t)dW(t)$$

for $F \in \mathbb{L}^1(0,T), G \in \mathbb{L}^2(0,T)$. Assume $u : \mathbb{R} \times [0,T] \to \mathbb{R}$ is continuous, and that

$$u_t = \frac{\partial u}{\partial t}, \qquad u_x = \frac{\partial u}{\partial x}, \qquad u_{xx} = \frac{\partial^2 u}{\partial x^2}$$

exist and are continuous. Then the stochastic process

$$Y(t) := u\big(X(t), t\big)$$

has the stochastic differential

$$dY = u_t dt + u_x dX + \frac{1}{2} u_{xx} G^2 dt \tag{69}$$

$$= \Big(u_t + u_x F + \frac{1}{2} u_{xx} G^2\Big) dt + u_x G dW$$

The unfamiliar term $\frac{1}{2} u_{xx} G^2 dt$ in (69) is a consequence of the *quadratic variation* of stochastic processes. For a smooth function $x(t)$, the basic relation between differentiation and integration can be stated as

$$x(t) = x(0) + \int_0^t \dot{x}(s) ds$$

or written in differential form

$$dx(t) = \dot{x}(t) dt$$

For $F \in C^2(\mathbb{R})$, Taylor's theorem states

$$\Delta F\big(x(t)\big) = F\big(x(t + \Delta t)\big) - F\big(x(t)\big)$$

$$= F'\big(x(t)\big) \Delta x(t) + \frac{1}{2} F''\big(x(t)\big) (\Delta x(t))^2$$

where $\Delta x(t) = x(t + \Delta t) - x(t)$. Taking the limit $\Delta t \to 0$ gives

$$dF\big(x(t)\big) = F'\big(x(t)\big)\Delta x(t)$$

since for a smooth function $x(t)$, as $\Delta t \to 0$,

$$\Delta x(t) \to dx(t) = \dot{x}(t)dt$$

and terms of higher order $(dt)^2$ disappear.

However, this classical relation no longer holds for stochastic processes since they are *nowhere differentiable*, and of *unbounded variation* in every given interval. Stochastic calculus can also be considered as an extension of classical calculus to functions of unbounded variation. Simply, when forming the differential $dF\big(X(t)\big)$ the second term of the Taylor expansion can no longer be neglected, since the term $(\Delta X(t))^2$, the quadratic variation of $X(t)$, does not disappear for $\Delta t \to 0$. Thus for functions of unbounded variation the differential is of the form

$$dF\big(X(t)\big) = F'\big(X(t)\big)dX(t) + \frac{1}{2}F''\big(X(t)\big)\big(dX(t)\big)^2$$

Example A.22 (sde solution)

$$\left\{ \begin{array}{c} dX = X dW \\ \\ X(0) = 1 \end{array} \right\} \tag{70}$$

has solution

$$X(t) = e^{W(t) - \dfrac{t}{2}}$$

and not, as one might incorrectly guess

$$X(t) = e^{W(t)}$$

In stochastic calculus, the expression

$$e^{\lambda W(t) - \dfrac{\lambda^2 t}{2}}$$

plays the role that $e^{\lambda t}$ *plays in deterministic calculus.*

Properties of the Itô integral include

$$\int (aG + bH) dW = a \int G dW + b \int H dW$$

$$\int_a^c G dW = \int_a^b G dW + \int_b^c G dW$$

$$d(W^2) = 2W\,dW + dt \tag{71}$$

$$d(tW) = W\,dt + t\,dW \tag{72}$$

$$\begin{aligned} \mathbb{E}\Big[\Big(\int_0^T G\,dW\Big)\Big(\int_0^T H\,dW\Big)\Big] &= \mathbb{E}\Big[\int_0^T GH\,dW\Big] \\ \Rightarrow \quad \mathbb{E}\Big[\Big(\int_0^T G\,dW\Big)^2\Big] &= \mathbb{E}\Big[\int_0^T G^2\,dW\Big] \end{aligned} \tag{73}$$

where $\mathbb{E}$ denotes expectation. (71) and (72) are special cases of the following product rule for stochastic processes. (73) is known as the *Itô isometry*.

Theorem A.23 (Itô's product rule / integration-by-parts formula) *Suppose*

$$\left\{ \begin{aligned} dU &= F_1\,dt + G_1\,dW \\ dV &= F_2\,dt + G_2\,dW \end{aligned} \right\}, \qquad (0 \le t \le T)$$

for $F_1, F_2 \in \mathbb{L}^1(0,T)$ and $G_1, G_2 \in \mathbb{L}^2(0,T)$. Then

$$d(UV) = U\,dV + V\,dU + G_1 G_2\,dt$$

and the integrated version of this product rule is known as the Itô integration-by-parts formula

$$\int_r^s U\,dV = UV\Big|_r^s - \int_r^s V\,dU - \int_r^s G_1 G_2\,dt$$

As with most other types of differential equations, *"[analytically] solvable SDEs are rare in practical applications"* (Kloeden and Platen, 1992, p.xxi). Numerical methods for solving SDEs (Rumelin, 1982; Greiner et al., 1988; Kloeden and Platen, 1989; Talay and Tubaro, 1990; Kloeden and Platen, 1992; Saito and Mitsui, 1996; Artemiev and Averina, 1997; Mauthner, 1998; Platen et al., 1999; Burrage et al., 2000) typically *"involve the simulation of a large number of different sample paths in order to estimate various statistical features of the desired solution"* (Kloeden and Platen, 1992, p.xxi). Modern parallel architecture supercomputers are well-suited to such calculations (Klauder and Petersen, 1985; Petersen, 1987, 1990).

The simplest discrete time approximation of SDEs is the stochastic generalization of Euler's method, known as the *Euler-Maruyama approximation* (Maruyama, 1955; Yuan and Mao, 2004; Lamba et al., 2007). For a time discretization of an interval $[0, T]$,

$$0 = \tau_0 < \tau_1 < \cdots < \tau_n < \cdots < \tau_N = T$$

and initial value $Y_0 = x_0$, the Euler-Maruyama approximation has the form

$$Y_{n+1} = Y_n + f(Y_n)\Delta_n + g(Y_n)\Delta W_n, \qquad n = 0, 1, \ldots, N-1$$

where

$$\Delta_n = \tau_{n+1} - \tau_n, \qquad \Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$$

The random variables ΔW_n are independent $N(0, \Delta_n)$ normally-distributed random variables, which can be provided by any standard pseudo-random number generator.

One important application of SDEs is the stochastic representation for solutions to PDEs, known as the *Feynmann-Kac formula* (Mao, 1997, p. 78-84). This formula is a link between SDEs and PDEs, and creates the probabilistic approach to the study of partial differential equations (Friedlin, 1985; Graham et al., 1996; Cerrai, 2001). It offers a method of solving certain PDEs by simulating random paths of a stochastic process.

Another important application of SDEs is the Black-Scholes (1973) option pricing formula. Assume a stock price $Y(t)$ can be described by the stochastic differential equation

$$dY = \mu Y\,dt + \sigma Y\,dW$$

where μ is the expected return per unit time, and σ^2 is the variance per unit time. This geometric Brownian motion model (Figure 45) implies that the changes in the logarithm of the price are normally distributed, ie.

$$\log \dot{Y} \sim N(\mu, \sigma^2)$$

Financial derivatives, of which options are a special case, are so called because they "derive" their value from some other security. Itô's lemma is the stochastic version of the chain rule. Therefore any function of the price Y, including the price of an option $C(Y)$, must satisfy (69) with $F = \mu Y$ and $G = \sigma Y$,

$$dC = \left[C_t + C_Y \mu Y + \frac{1}{2} C_{YY} \sigma^2 Y^2\right] dt + C_Y \sigma Y\, dW$$

where

$$C_t = \frac{\partial C}{\partial t}, \qquad C_Y = \frac{\partial C}{\partial Y}, \qquad C_{YY} = \frac{\partial^2 C}{\partial Y^2}$$

The key insight of Black-Scholes is that, under idealized conditions such as trading possible in continuous time with no transaction costs, two opposite risky positions taken together can effectively eliminate risk itself. Consider a riskless portfolio of a time-varying number of shares of the underlying stock and a short option position

$$\phi = Y C_Y - C$$

where C_Y represents the continuous adjusting of this riskless portfolio over time by buying or selling shares of the underlying stock to stay perfectly hedged, something known as *dynamical hedging*. The change in the value of this portfolio is

$$
\begin{aligned}
d\phi &= dY C_Y - dC \\
&= \left(\cancel{\mu Y dt} + \cancel{\sigma Y dW}\right) C_Y - \left[C_t + \cancel{C_Y \mu Y} + \frac{1}{2} C_{YY} \sigma^2 Y^2\right] dt - \cancel{C_Y \sigma Y dW} \\
&= -\left(C_t + \frac{1}{2} C_{YY} \sigma^2 Y^2\right) dt \qquad (74)
\end{aligned}
$$

In the absence of arbitrage opportunities, $d\phi$ must also equal the gain obtained by investing the same amount of money in a security paying the constant riskless rate r,

$$d\phi = r\phi dt = (rY C_Y - rC) dt \qquad (75)$$

By equating (74) and (75), we obtain

$$C_t + rY C_Y + \frac{1}{2} C_{YY} \sigma^2 Y^2 = rC \qquad (76)$$

which is known as the *Black-Scholes-Merton SPDE*, subject to boundary conditions

$$C(T) = (Y(T) - K)^+ = \max(Y(T) - K, 0)$$

$$P(T) = (K - Y(T))^+ = \max(K - Y(T), 0)$$

for call and put options respectively. Making the substitutions

$$C(Y, t) = e^{r(t-T)} y(x, s)$$

where

$$x = \frac{2}{\sigma^2}\big(r - \frac{\sigma^2}{2}\big)\Big[\log\big(\frac{Y}{K}\big) - \big(r - \frac{\sigma^2}{2}\big)(t - T)\Big]$$

and

$$s = -\frac{2}{\sigma^2}\big(r - \frac{\sigma^2}{2}\big)(t - T)$$

gives the heat equation

$$y_s = y_{xx}$$

which can be solved analytically to obtain the *Black-Scholes option formula*

$$C(Y, t) = Y N(d_1) - K e^{r(t-T)} N(d_2) \tag{77}$$

where

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy,$$

$$d_1 = \frac{\log(Y/K) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}},$$

$$d_2 = d_1 - \sigma\sqrt{T - t}.$$

There are several extensions to Black-Scholes that relax various assumptions which might not be realistic in actual financial markets. Examples include option pricing with a stochastic interest rate (Merton, 1973; Amin and Jarrow, 1992), jump processes (Merton, 1976), jump-diffusion processes (Bates, 1991), stochastic volatility (Hull and White, 1987; Heston, 1993), and non-gaussian distributions (Bouchaud and Sornette, 1994; Aurell et al., 2000; Matacz, 2000).

A.3.3 Stochastic Delay Differential Equations

Literature on stochastic delay differential equations is limited to few authors. Mohammed (1996); Mohammed and Scheutzow (1997); Mohammed (2003) and Mao (2003a,b) pioneered this field with the first articles and monographs on the subject. Starting with his PhD thesis, Reiss (2002, 2003, 2005); Reiss et al. (2006) considered the inverse problem for SDDEs. Buckwar (2000, 2006); Kuchler and Platen (2000); Küchler and Platen (2002); Küchler and Vasiliev (2005); Küchler and Sørensen (2006) published a series of articles on numerical methods for SDDEs. Stoica (2005); Arriojas et al. (2007a,b); Küchler and Platen (2007) only recently started looking at finance applications of SDDEs.

Definition A.24 (stochastic delay differential equation (SDDE))

$$dX(t) = f(t, X_t)\ dt + g(t, X_t)\ dB(t) \tag{78}$$

This definition of an SDDE is quite general and incorporates most other differential equations as special cases (Table 8).

MODEL	**SDDE** $dX(t) = f(t, X_t)\ dt + g(t, X_t)\ dB(t)$
Brownian Motion	$dX(t) = \mu\ dt + \sigma\ dB(t)$
Geometric Brownian Motion	$dX(t) = \mu X(t)\ dt + \sigma X(t)\ dB(t)$
Ornstein-Uhlenbeck	$dX(t) = \theta\big(\mu - X(t)\big)\ dt + \sigma X(t)\ dB(t)$
ODE	$dX(t) = f\big(t, X(t)\big)\ dt$
DDE	$dX(t) = f(t, X_t)\ dt$
AR(p)	$dX(t) = f(X_t)\ dt + \sigma\ dB(t),\quad$ where $f(X_t) = f\big(X(t-1), X(t-2), \ldots, X(t-p)\big)$
ARCH(p,q)	$dX(t) = f(X_t)\ dt + \sigma(X_t)\ dB(t),\quad$ where $f(X_t) = f\big(X(t-1), X(t-2), \ldots, X(t-p)\big)$ $\sigma(X_t) = Var\big(X(t-1), X(t-2), \ldots, X(t-q)\big)$

Table 8: SDDE Models

Ornstein-Uhlenbeck stochastic processes are "mean-reverting", a continuous version of discrete autoregressive moving average (ARMA). Deterministic differential equations, and delay differential equations, are also a special case of (78) where $g(t, X_t) = 0$, ie.

$$dX(t) = f(t, X_t)\ dt \quad \Rightarrow \quad \frac{dX(t)}{dt} = f(t, X_t)$$

In econometrics, autoregressive AR(p) models are discrete time series models, where the current rate of change is a function of the recent past. In the case of linear dependence,

$$f(X_t) = \beta_0 + \beta_1 X(t-1) + \ldots + \beta_p X(t-p)$$

Murphy (2006) discussed nonlinear polynomial autoregressive models PAR(p,n), ie.

$$f(X_t) = \sum \beta_i P_n\big(X(t-1), \ldots, X(t-p)\big)$$

where P_n are multidimensional Legendre orthogonal polynomials of degree n.

Autoregressive conditional heteroskedasticity (ARCH, (Engle, 1982; Bollerslev et al., 1992; Greene, 2003; Kennedy, 2003)) models are an extension of brownian motion that accounts for serial correlation in volatility, where the present variance is a function of the variance in the recent past.

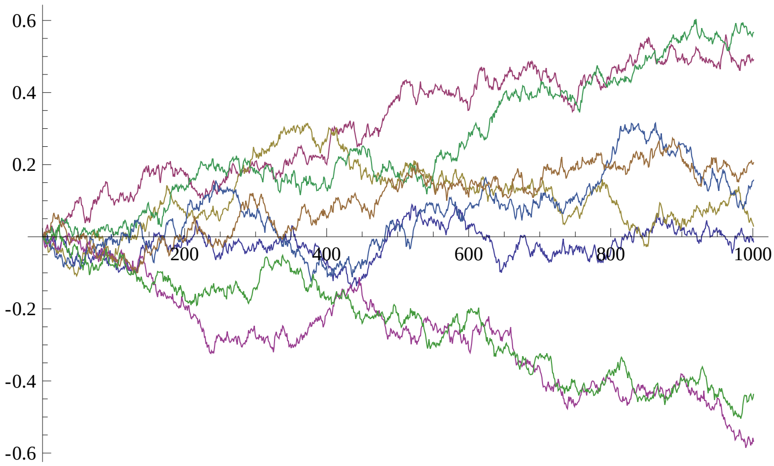


Figure 44: Brownian Motion

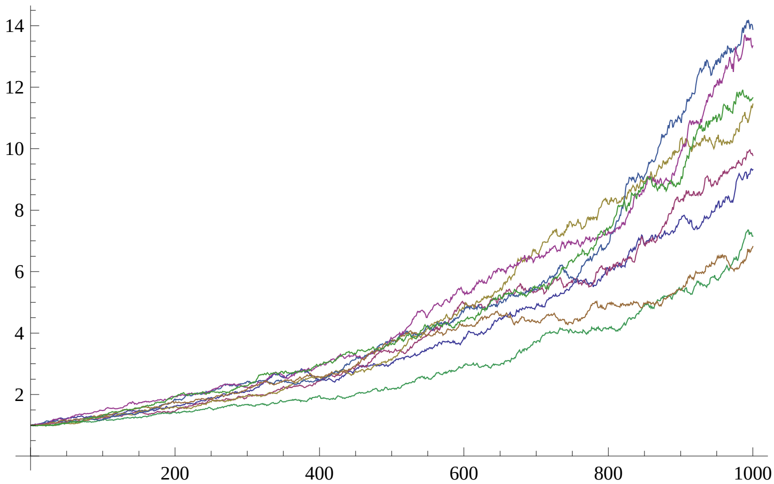


Figure 45: Geometric Brownian Motion

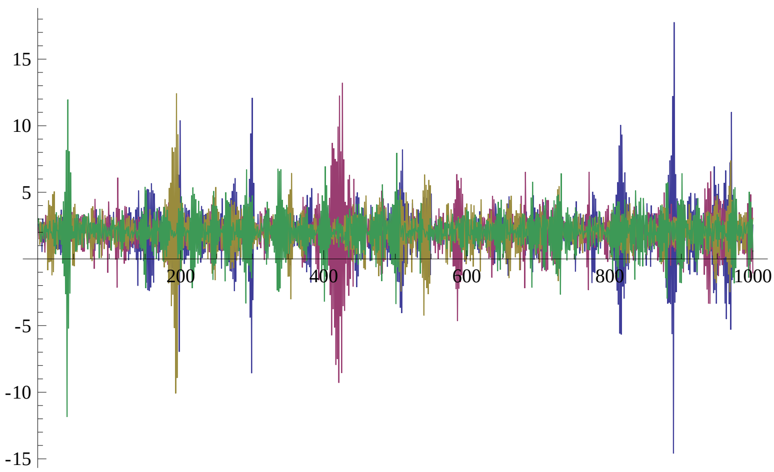


Figure 46: Ornstein-Uhlenbeck Process

References

Aliprantis, C. and Border, K. (1999). *Infinite Dimensional Analysis: a hitchhiker's guide*. Springer. 114

Alpert, M. and Raiffa, H. (1982). A progress report on the training of probability assessors. In *Judgment Under Uncertainty: Heuristics and Biases*, pages 294–305. Cambridge University Press. 98

Amin, K. and Jarrow, R. (1992). Pricing options on risky assets in a stochastic interest rate economy. *Mathematical Finance*, 2(4):217–237. 135

Arino, O., Hbid, M., and Dads, E. (2006). *Delay Differential Equations and Applications*. Springer. 114

Arriojas, M., Hu, Y., Mohammed, S., and Pap, G. (2007a). A Delayed Black and Scholes Formula I. *Arxiv preprint math.PR/0604640*. 136

Arriojas, M., Hu, Y., Mohammed, S., and Pap, G. (2007b). A Delayed Black and Scholes Formula II. *Arxiv preprint math.PR/0604641*. 136

Artemiev, S. and Averina, T. (1997). *Numerical Analysis of Systems of Ordinary and Stochastic Differential Equations*. Vsp. 130

Aurell, E., Baviera, R., Hammarlid, O., Serva, M., and Vulpiani, A. (2000). A general methodology to price and hedge derivatives in incomplete markets. *International Journal of Theoretical and Applied Finance*, 3(1):1–24. 135

Azzalini, A. and Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726. 5

Bachelier, L. (1900). Theorie de la speculation. *Annales Scientifiques de l'Ecole Normale Superieure*, 3(17):21–86. 59, 93

Barberis, N., Huang, M., and Santos, T. (2001). Prospect Theory and Asset Prices. *The Quarterly Journal of Economics*, 116(1):1–53. 101

Barberis, N. and Thaler, R. (2002). A Survey of Behavioral Finance. 96, 100

Bates, D. (1991). The Crash of'87: Was It Expected? The Evidence from Options Markets. *The Journal of Finance*, 46(3):1009–1044. 135

Bell, D. (1982). Regret in Decision Making under Uncertainty. *Operations Research*, 30(5):961–981. 101

Bellen, A. and Zennaro, M. (2003). *Numerical Methods for Delay Differential Equations*. Oxford University Press. 114

Benartzi, S. and Thaler, R. (1995). Myopic Loss Aversion and the Equity Premium Puzzle. *The Quarterly Journal of Economics*, 110(1):73–92. 101, 104

Berntsen, J., Espelid, T., and Genz, A. (1991). An adaptive algorithm for the approximate calculation of multiple integrals. *ACM Transactions on Mathematical Software (TOMS)*, 17(4):437–451. 9, 40

Black, F. (1986). Noise. *The Journal of Finance*, 41(3):529–543. 95

Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, 81(3):637–654. 59, 131

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992). ARCH modeling in finance : A review of the theory and empirical evidence. *Journal of Econometrics*, 52(1):5–59. 137

Bookstaber, R. and Clarke, R. (1984). Option portfolio strategies: measurement and evaluation. *Journal of Business*, 57(4):469–492. 65

Bouchaud, J. and Sornette, D. (1994). The Black-Scholes option pricing problem in mathematical finance: generalization and extensions for a large class of stochastic processes. *J. Phys. I France*, 4:863–881. 135

Bowman, K. and Shenton, L. (1979a). Approximate percentage points for Pearson distributions. *Biometrika*, 66(1):147–151. 57

Bowman, K. and Shenton, L. (1979b). Further approximate Pearson percentage points and Cornish-Fisher. *Communications in Statistics-Simulation and Computation*, 8(3):231–244. 57

Bratley, P. and Fox, B. (1988). ALGORITHM 659: implementing Sobol's quasirandom sequence generator. *ACM Transactions on Mathematical Software (TOMS)*, 14(1):88–100. 40, 41

Brennan, M., Jegadeesh, N., and Swaminathan, B. (1993). Investment Analysis and the Adjustment of Stock Prices to Common Information. *The Review of Financial Studies*, 6(4):799–824. 106

Brown, A. (2002). The unbearable lightness of cross-market risk. *Wilmott*, 2002(1):20–23. 56

Brown, S., Goetzmann, W., and Ross, S. (1995). Survival. *The Journal of Finance*, 50(3):853–873. 104

Buckwar, E. (2000). Introduction to the numerical analysis of stochastic delay differential equations. *Journal of Computational and Applied Mathematics*, 125(1-2):297–307. 135

Buckwar, E. (2006). One-step approximations for stochastic functional differential equations. *Applied Numerical Mathematics*, 56(5):667–681. 135

Buehler, R., Griffin, D., and Ross, M. (1994). Exploring the planning fallacy: why people underestimate their task completion times. *Journal of personality and social psychology*, 67(3):366–381. 99

Burrage, K., Burrage, P., and Mitsui, T. (2000). Numerical solutions of stochastic differential equations–implementation and stability issues. *Journal of Computational and Applied Mathematics*, 125(1-2):171–182. 130

Campbell, J. Y. (1999). Asset prices, consumption, and the business cycle. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*. Elsevier. 104

Casella, G. and Berger, R. (2002). *Statistical inference*. Duxbury. 48

Catmull, E. and Clark, J. (1978). Recursively generated B-spline surfaces on arbitrary topological meshes. *Computer-Aided Design*, 10(6):350–355. 54

Cerrai, S. (2001). *Second order PDE's in finite and infinite dimension: a probabilistic approach*. Lecture notes in mathematics 1762. Springer. 131

Chen, H., Jegadeesh, N., and Wermers, R. (2000). The Value of Active Mutual Fund Management: An Examination of the Stockholdings and Trades of Fund Managers. *The Journal of Financial and Quantitative Analysis*, 35(3):343–368. 105

Cherney, A. and Engelbert, H. (2005). *Singular Stochastic Differential Equations*. Springer. 122

Chew, S. (1989). Axiomatic utility theories with the betweenness property. *Annals of Operations Research*, 19(1):273–298. 101

Chew, S. and MacCrimmon, K. (1979). *Alpha-Nu Choice Theory: A Generalization of Expected Utility Theory*. The University of British Columbia. 101

Clark, M. W. (2000). Trillion dollar bet [videotape]. *WGBH/PBS NOVA Series*. 93

Cornish, E. and Fisher, R. (1938). Moments and cumulants in the specification of distributions. *Revue de l'Institut International de Statistique*, pages 307–320. 57

Cox, J. C. and Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3(1-2):145–166. available at http://ideas.repec.org/a/eee/jfinec/v3y1976i1-2p145-166.html. 59

Da Prato, G. (2006). *An Introduction to Infinite-dimensional Analysis*. Springer. 114

Da Prato, G. and Tubaro, L. (2002). *Stochastic Partial Differential Equations and Applications*. Marcel Dekker. 122

Da Prato, G. and Zabczyk, J. (1992). *Stochastic equations in infinite dimensions*. Cambridge University Press New York. 122

Davenport, J. and Herring, T. (1979). On the use of curve fitting to model the error of the Cornish-Fisher expansion of the pearson type-IV distribution. *Communications in Statistics-Simulation and Computation*, 8(4):311–333. 57

Davies, J., Grant, M., and Venezia, J. (2006). Transportation greenhouse gas emissions: Trends, uncertainties and methodological improvements. In *15th International Emission Inventory Conference*. 73

de Doncker, E., Cucos, L., Zanny, R., and Genz, A. (2001). Parallel computation of the multivariate t-distribution. In *Proceedings of the High Performance Computing Symposium 2001 (HPC'01)*. 46

de Doncker, E., Genz, A., and Ciobanu, M. (1999). Parallel computation of multivariate normal probabilities. *Computing Science and Statistics*, 30:89–93. 46

de Doncker, E., Kaugars, K., Cucos, L., and Zanny, R. (2002). Current status of the ParInt package for parallel multivariate integration. In *KEK PROCEEDINGS*, pages 110–119. High Energy Accelerator Research Organization; 1999. 46

de Long, J., Shleifer, A., Summers, L., and Waldmann, R. (1990a). Noise Trader Risk in Financial Markets. *The Journal of Political Economy*, 98(4):703–738. 95

de Long, J., Shleifer, A., Summers, L., and Waldmann, R. (1990b). Positive Feedback Investment Strategies and Destabilizing Rational Speculation. *The Journal of Finance*, 45(2):379–395. 96

Dedoncker, E., Gupta, A., Ball, J., Ealy, P., and Genz, A. (1996). Parint: A software package for parallel integration. In *Proceedings of the 10th international conference on Supercomputing*, pages 149–156. ACM. 46

Dekel, E. (1986). An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom. *Journal of Economic Theory*, 40(2):304–18. 101

Elderton, W. and Johnson, N. (1969). *Systems of frequency curves*. Cambridge University Press London. vii, 18

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, 75(4):643–669. 104

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, 50:987–1008. 137

Evans, L. (2006). An introduction to stochastic differential equations. Lecture Notes `http://math.berkeley.edu/~evans/SDE.course.pdf`. 122

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417. 93

Fang, H.-B., Fang, K.-T., and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82(1):1 – 16. 1

Farmer, J. (2002). Market force, ecology and evolution. *Industrial and Corporate Change*, 11(5):895–953. v, 106, 109, 110, 111

Finney, D. (1963). Some properties of a distribution specified by its cumulants. *Technometrics*, 5(1):63–69. 57

Fischhoff, B., Slovic, P., and Lichtenstein, S. (1977). Knowing with Certainty: The Appropriateness of Extreme Confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3:552–564. 98

Fisher, R. and Cornish, E. (1960). The percentile points of distributions having known cumulants. *Technometrics*, 2(2):209–225. 57

Forman, J. and Sørensen, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian journal of statistics*, 35(3):438–465. 78

Frey, C. (2007). Quantification of uncertainty in emission factors and inventories. In *16th Annual International Emission Inventory Conference*. vii, 72

Frey, C., Penman, J., Hanle, L., Monni, S., and Ogle, S. (2006). *IPCC Guidelines for National Greenhouse Gas Inventories*, volume 1 General Guidance and Reporting, chapter 3 Uncertainties, pages 3.1 – 3.66. IGES. 72

Friedlin, M. (1985). Functional Integration and Partial Differential Equations. *Ann. of Math. Studies*, 109. 131

Friedman, A. (1975). *Stochastic Differential Equations and Their Applications, Volume 1*. Academic Press. 122

Friedman, A. (1976). Stochastic Differential Equations and Their Applications, Volume 2. 122

Friedman, J. and Wright, M. (1981). A nested partitioning procedure for numerical multiple integration. *ACM Transactions on Mathematical Software (TOMS)*, 7(1):76–92. 41

Froot, K. and Dabora, E. (1999). How are stock prices affected by the location of trade? *Journal of Financial Economics*, 53(189):216. vii, 97

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M., and Rossi, F. (2010). *GNU Scientific Library Reference Manual (v1.14)*. 48

Gassmann, H., Deák, I., and Szántai, T. (2002). Computing Multivariate Normal Probabilities. *Journal of Computational and Graphical Statistics*, 11(4):920–949. 40

Genz, A. and Bretz, F. (2002). Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, 11(4):950–971. 40

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Number 195 in Lecture Notes in Statistics. Springer-Verlag. 46, 87

Genz, A. and Cools, R. (2003). An adaptive numerical cubature algorithm for simplices. *ACM Transactions on Mathematical Software (TOMS)*, 29(3):297–308. 9, 40

Genz, A. and Kass, R. (1997). Subregion-adaptive integration of functions having a dominant peak. *Journal of Computational and Graphical Statistics*, 6(1):92–111. 35

Genz, A. and Malik, A. (1980a). An adaptive algorithm for numerical integration over an N-dimensional rectangular region. *J. Comput. Appl. Math*, 6(4):295–302. 9, 40

Genz, A. and Malik, A. (1980b). Remarks on algorithm 006: An adaptive algorithm for numerical integration over an N-dimensional rectangular region. *Journal of Computational and Applied mathematics*, 6(4):295–302. 9, 40

Genz, A. and Malik, A. (1983). An imbedded family of fully symmetric numerical integration rules. *SIAM Journal on Numerical Analysis*, 20(3):580–588. 40, 41

Gervais, S. and Odean, T. (2001). Learning to Be Overconfident. *The Review of Financial Studies*, 14(1):1–27. 98

Gilovich, T., Griffin, D., Griffin, D., and Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press. 98

Gosling, J. and O'Hagan, A. (2007). Understanding the uncertainty in the biospheric carbon flux for England and Wales. 73

Graham, C. et al. (1996). *Probabilistic models for nonlinear partial differential equations*. Lecture notes in mathematics 1627. Springer. 131

Greene, W. (2003). *Econometric Analysis (5th Ed.)*. Prentice Hall. 137

Greiner, A., Strittmatter, W., and Honerkamp, J. (1988). Numerical integration of stochastic differential equations. *Journal of Statistical Physics*, 51(1):95–108. 130

Griffin, J., Ji, X., and Martin, J. (2003). Momentum Investing and Business Cycle Risk: Evidence from Pole to Pole. *The Journal of Finance*, 58(6):2515–2547. 105

Grinblatt, M., Titman, S., and Wermers, R. (1995). Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior. *The American Economic Review*, 85(5):1088–1105. 105

Gu, A. (2003). The declining january effect: evidences from the u.s. equity markets. *The Quarterly Review of Economics and Finance*, 43(2):395–404. 106

Gul, F. (1991). A Theory of Disappointment Aversion. *Econometrica*, 59(3):667–686. 101

Gupta, A. (2003). Multivariate skew t-distribution. *Statistics*, 37(4):1–1. 5

Hahn, T. (2005). CUBA – a library for multidimensional numerical integration. *Computer physics communications*, 168(2):78–95. 9, 40, 46

Hajivassiliou, V. (1993). Simulating normal rectangle probabilities and their derivatives: Effects of vectorization. *International Journal of High Performance Computing Applications*, 7(3):231. 46

Hale, J., Lunel, V., Sjoerd, M., Verduyn, L., Lunel, V., and Sjoerd, M. (1993). *Introduction to Functional Differential Equations*. Springer Verlag. 114, 119

Hansen, L. and Singleton, K. (1983). Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns. *The Journal of Political Economy*, 91(2):249–265. 104

Heston, S. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343. 135

Hong, C. (1983). A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox. *Econometrica*, 51(4):1065–1092. 101

Hong, H. and Hickernell, F. (2003). Algorithm 823: Implementing scrambled digital sequences. *ACM Transactions on Mathematical Software (TOMS)*, 29(2):109. 40

Hull, J. and White, A. (1987). The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance*, 42(2):281–300. 135

Ibragimov, I., Linnik, I., and Kingman, J. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff. 90

ICF (2004). Quantitative assessment of uncertainty in Canada's national GHG inventory estimates for 2001. 73

ICF (2005). Quantitative assessment of uncertainty in Canada's national GHG inventory estimates for 2001: Supplementary analysis. 73

Ito, K. (1944). Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20(519-524). 59, 125

Ito, K. (1946). On a stochastic integral equation. *Proc. Japan Acad*, 22:32–35. 59, 125

Jackson, P., Maude, D., Perraudin, W., and England, B. (1998). Bank capital and Value at Risk. *Journal of Derivatives*, 4(3):73–90. 55

Jegadeesh, N., Kim, J., Krische, S., and Lee, C. (2004). Analyzing the Analysts: When Do Recommendations Add Value? *The Journal of Finance*, 59(3):1083–1124. 105

Jegadeesh, N. and Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1):65–91. 105, 106

Jegadeesh, N. and Titman, S. (1995). Overreaction, Delayed Reaction, and Contrarian Profits. *The Review of Financial Studies*, 8(4):973–993. 106

Johnson, N., Kotz, S., and Balakrishnan, N. (2002). *Continuous Multivariate Distributions, volume 1, Models and Applications*. New York: John Wiley & Sons,. 16

Jones, M. (2002). Marginal replacement in multivariate densities, with application to skewing spherically symmetric distributions. *Journal of Multivariate Analysis*, 81(1):85–99. 6

Jones, M. and Faddy, M. (2003). A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):159–174. 6

Jorion, P. (2007). *Value at risk: the new benchmark for managing financial risk*. McGraw-Hill Companies. 57

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press. 98

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292. vii, 101, 102

Kahneman, D. and Tversky, A. (2000). *Choices, Values, and Frames*. Cambridge University Press. 98

Kang, H. (1997). 3D lookup table with interpolation. *Color Technology of Electronic Imaging Devices*, 28. vi, 51, 52

Keng, H. and Yuan, W. (1981). *Applications of number theory to numerical analysis*. Springer-Verlag, Berlin; New York. 41

Kennedy, M., Anderson, C., O'Hagan, A., Lomas, M., Woodward, I., Gosling, J. P., and Heinemeyer, A. (2008). Quantifying uncertainty in the biospheric carbon flux for england and wales. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(1):109–135. 73

Kennedy, P. (2003). *A Guide to Econometrics (5th Ed.)*. Blackwell Publishing. 137

Klauder, J. and Petersen, W. (1985). Numerical Integration of Multiplicative-Noise Stochastic Differential Equations. *SIAM Journal on Numerical Analysis*, 22(6):1153–1166. 130

Kloeden, P. and Platen, E. (1989). A survey of numerical methods for stochastic differential equations. *Stochastic Environmental Research and Risk Assessment (SERRA)*, 3(3):155–178. 130

Kloeden, P. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer. 122, 130

Knight, F. (1921). *Risk, Uncertainty and Profit*. Houghton Mifflin Company. 104

Kolman, J., Onak, M., Jorion, P., Taleb, N., Derman, E., Putnam, B., Sandor, R., Jonas, S., Dembo, R., Holt, G., et al. (1997). Roundtable: The Limits of VaR. *Derivatives Strategy*, 3:14–22. 57

Kolmanovskii, V. and Myshkis, A. (1999). *Introduction to the theory and applications of functional differential equations*. Kluwer Academic Publishers Boston. 113, 114

Korobov, N. (1963). Number Theoretic Methods in Approximate Analysis. 41

Kotz, S. and Nadarajah, S. (2004). *Multivariate t distributions and their applications*. Cambridge Univ Press. 9, 32, 87

Kotz, S. and Vicari, D. (2005). Survey of developments in the theory of continuous skewed distributions. *Metron*, 53:225–261. 5

Kuchler, U. and Platen, E. (2000). Strong discrete time approximation of stochastic differential equations with time delay. *Mathematics and Computers in Simulation*, 54(1):189–205. 135

Küchler, U. and Platen, E. (2002). Weak discrete time approximation of stochastic differential equations with time delay. *Mathematics and Computers in Simulation*, 59(6):497–507. 135

Küchler, U. and Platen, E. (2007). Time delay and noise explaining cyclical fluctuations in prices of commodities. available at http://ideas.repec.org/p/uts/rpaper/195.html. 79, 136

Küchler, U. and Sørensen, M. (2006). Statistical inference for discrete-time samples from affine stochastic delay differential equations. 135

Küchler, U. and Vasiliev, V. (2005). Sequential Identification of Linear Dynamic Systems with Memory. *Statistical Inference for Stochastic Processes*, 8(1):1–24. 135

Lalescu, C. et al. (2009). Two hierarchies of spline interpolations. Practical algorithms for multivariate higher order splines. *Arxiv preprint arXiv:0905.3564*. 54

Lalescu, C., Teaca, B., and Carati, D. (2010). Implementation of high order spline interpolations for tracking test particles in discretized fields. *Journal of computational physics*, 229(17):5862–5869. 54

Lamba, H., Mattingly, J., and Stuart, A. (2007). An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *IMA Journal of Numerical Analysis*, 27(3):479. 130

Lepage, G. (1980). Vegas: An adaptive multidimensional integration program. *Cornell Preprint, CLNS*. 40

Lepage, P. et al. (1978). A new algorithm for adaptive multidimensional integration* 1. *Journal of Computational Physics*, 27(2):192–203. 40

LeRoy, S. and Porter, R. (1981). The Present-Value Relation: Tests Based on Implied Variance Bounds. *Econometrica*, 49(3):555–574. 104

Lo, A. and MacKinlay, A. (1990). When are contrarian profits due to market overreaction. *Review of Financial Studies*, 3(2):175–205. 106

Loh, W. (2004). On the characteristic function of Pearson type IV distributions. *Lecture Notes-Monograph Series*, 45:171–179. 89

Loomes, G. and Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, 92(368):805–824. 101

Lord, C., Ross, L., and Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11):2098–2109. 100

Magalhäaes, L., Oliva, W., and Hale, J. (2002). *Dynamics in Infinite Dimensions*. Springer. 114

Mao, X. (1997). *Stochastic differential equations and applications*. Horwood. 131

Mao, X. (2003a). Numerical solutions of stochastic functional differential equations. *LMS J. Comput. Math*, 6(2). 135

Mao, X. (2003b). Stochastic differential delay equations. *Lecture notes for Spring School in Stochastic Delay Differential Equations, Humboldt University Berlin, March*. 135

Maruyama, G. (1955). Continuous Markov processes and stochastic equations. *Rend. Circ. Mat. Palermo II Ser*, 4:48–90. 130

Matacz, A. (2000). Financial modeling and option theory with the truncated levy process. *International Journal of Theoretical and Applied Finance*, 3(1):143–160. 135

Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30. 41

Mauthner, S. (1998). Step size control in the numerical solution of stochastic differential equations. *Journal of Computational and Applied Mathematics*, 100(1):93–109. 130

McCann (1994). Uncertainties in Canada's 1990 GHG emission estimates. 73

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools*. Princeton Univ Press. 55

Mehra, R. and Prescott, E. (1985). The Equity Premium: A Puzzle. *Journal of Monetary Economics*, 15(2):145–61. 104

Merton, R. (1973). Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183. 135

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144. 135

Merton, R. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance*, 42(3):483–510. 95

Merton, R., Scholes, M., and Gladstein, M. (1978). The returns and risk of alternative call option portfolio investment strategies. *Journal of Business*, 51(2):183–242. 65

Mitic, P. (1996). Critical values for the Wilcoxon signed rank statistic. *The Mathematica Journal*, 6(3):73–77. 63

Mohammed, S. (1996). Stochastic differential systems with memory: Theory, examples and applications. *Stochastic Analysis and Related Topics VI. The Geilo Workshop*, 42:1–77. 135

Mohammed, S. (2003). Stochastic differential systems with memory: Theory, examples and applications. *Berling Sprint School on Stochastic Delay Equations*, 42. 135

Mohammed, S. and Scheutzow, M. (1997). Lyapunov Exponents of Linear Stochastic Functional Differential Equations. Part II. Examples and Case Studies. *The Annals of Probability*, 25(3):1210–1240. 135

Murphy, M. (2006). Nonlinear financial time series analysis using multivariate orthogonal polynomials. Master's thesis, Sprott School of Business, Carleton University. 95, 137

Murphy, M., Blain, D., Khakwani, O., Morales, A., McGovern, M., Metsaranta, J., and McConkey, B. (2009). Uncertainty analysis of greenhouse gas emissions/removals from agriculture, land use, land use change, and forestry in canada. Technical report, Environment Canada K8A46-08-0044. vii, 72, 73, 74, 76

Nagahara, Y. (1999). The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters. *Statistics & Probability Letters*, 43(3):251–264. 2

Nagahara, Y. (2003). Non-Gaussian filter and smoother based on the Pearson distribution system. *Journal of Time Series Analysis*, 24(6):721–738. 2

Nagahara, Y. (2004). A method of simulating multivariate nonnormal distributions by the Pearson distribution system and estimation. *Computational Statistics and Data Analysis*, 47(1):1–29. 2

NARSTO (2005). Improving emission inventories for effective air quality management across North America. 73

Narumi, S. (1923). On the General Forms of Bivariate Frequency Distributions Which are Mathematically Possible When Regression and Variation are Subjected to Limiting Conditions. *Biometrika*, 15(1):77–88. 5

Nelder, J. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7(4):308. 48

Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*. Society for Industrial Mathematics. 40

Nolan, J. (2008). Stable distributions. `http://academic2.american.edu/~jpnolan/stable/stable.html`. 90

Norden, R. (1972). A survey of maximum likelihood estimation. *International Statistical Review/Revue Internationale de Statistique*, 40(3):329–354. 48

Oksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications (6th Ed.)*. Springer. 122

Ord, J. (1972). *Families of frequency distributions*. Griffin London. vii, 3

Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A*, pages 343–414. 2

Petersen, W. (1987). Numerical simulation of Itô stochastic differential equations on supercomputers. *Random media, IMA Vol. Math. Appl.*, 7:215–225. 130

Petersen, W. (1990). *Stability and Accuracy of Simulations for Stochastic Differential Equations*. ETH. 130

Piessens, R., de Doncker-Kapenga, E., Überhuber, C., and Kahaner, D. (1983). *QUADPACK. A subroutine package for automatic integration*. Springer Series in Computational Mathematics, 1. 9, 34

Platen, E., University of Technology, S. S. o. F., Economics, and Group, Q. F. R. (1999). *An Introduction to Numerical Methods for Stochastic Differential Equations*. School of Finance and Economics, University of Technology, Sydney. 130

Press, W. and Farrar, G. (1990). Recursive stratified sampling for multidimensional Monte Carlo integration. *Computers in Physics*, 4(2):190–195. 40

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1993). *Numerical recipes in Fortran; the art of scientific computing*. Cambridge University Press New York, NY, USA. 40

Protter, P. (2003). *Stochastic Integration and Differential Equations*. Springer Verlag. 122

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3(4):323–343. 101

Rabin, M. (2002). A Perspective on Psychology and Economics. *European Economic Review*, 46(4-5):657–85. 100

Reiss, M. (2002). *Nonparametric Estimation for Stochastic Delay Differential Equations*. PhD thesis, Humboldt–Universitaet zu Berlin. 135

Reiss, M. (2003). Stochastic Delay Differential Equations. Lecture Notes. 135

Reiss, M. (2005). Adaptive estimation for affine stochastic delay differential equations. *Bernoulli*, 11(1):67–102. 135

Reiss, M., Riedle, M., den Linden, U., and van Gaans, O. (2006). Delay differential equations driven by Levy processes: stationarity and Feller properties. *Stochastic Processes and their Applications*, 116(10):1409–1432. 135

Rhodes, E. (1923). On a certain skew correlation surface. *Biometrika*, 14(3):355–377. 5

Rhodes, E. (1925). On a Skew Correlation Surface. *Biometrika*, 17(3):314–326. 5

Rumelin, W. (1982). Numerical Treatment of Stochastic Differential Equations. *SIAM Journal on Numerical Analysis*, 19(3):604–613. 130

Sahu, S., Dey, D., and Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics*, 31(2):129–150. 6

Saito, Y. and Mitsui, T. (1996). Stability Analysis of Numerical Schemes for Stochastic Differential Equations. *SIAM Journal on Numerical Analysis*, 33(6):2254–2267. 130

Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6:41–49. 93

Segal, U. (1987). Some remarks on Quiggin's anticipated utility. *Journal of Economic Behavior and Organization*, 8:145–154. 101

Segal, U. (1989). Anticipated utility: A measure representation approach. *Annals of Operations Research*, 19(1):359–373. 101

SGA (2000). Emission factors and uncertainties for CH_4 & N_2O from fuel combustion. Technical report, Prepared by SGA Energy Limited for Greenhouse Gas Division. 73

Shampine, L. and Thompson, S. (2000). Solving delay differential equations with DDE23. *Manuscript, available at http://www. runet. edu/thompson/webddes/index. html*. 114

Shampine, L. and Thompson, S. (2001). Solving DDEs in Matlab. *Applied Numerical Mathematics*, 37(4):441–458. 114

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442. 93

Shiller, R. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *The American Economic Review*, 71(3):421–436. 104

Shleifer, A. and Summers, L. (1990). The Noise Trader Approach to Finance. *The Journal of Economic Perspectives*, 4(2):19–33. 95

Shleifer, A. and Vishny, R. (1997). The Limits of Arbitrage. *The Journal of Finance*, 52(1):35–55. 95

Siegel, S. (1956). *Nonparametric methods for the behavioral sciences*. McGraw-Hill. 63

Sørensen, M. (2009). Parametric inference for discretely sampled stochastic differential equations. *Handbook of Financial Time Series*, pages 531–553. 78

Stein, C. (1986). Approximate computation of expectations. In *IMS Lecture Notes - Monograph Series*, volume 7. 89

Stoica, G. (2005). A stochastic delay financial model. *Proceedings of the American Mathematical Society*, 133(6):1837–1841. 136

Stuart, A. and Ord, J. (1994). *Kendall's advanced theory of statistics. Vol. 1: Distribution theory*. Hodder Arnold. 16, 48

Talay, D. and Tubaro, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4):483–509. 130

Thaler, R. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, 12(3):183–206. 104

Torrilhon, M. (2010). Hyperbolic Moment Equations in Kinetic Gas Theory Based on Multi-Variate Pearson-IV-Distributions. *Communications in Computational Physics*, 7(4):639–673. 7

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124. 99

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323. 101, 102, 103

Van Uven, M. (1925). On treating skew correlation I-II. *Proc. Roy. Acad. Sci. Amsterdam*, 28:797–811, 919–935. 5, 17

Van Uven, M. (1926). On treating skew correlation III. *Proc. Roy. Acad. Sci. Amsterdam*, 29:580–590. 5, 17

Van Uven, M. (1929). Skew correlation between three and more variables, I-III. *Proc. Roy. Acad. Sci. Amsterdam*, 32:793–807. 5, 17

Van Uven, M. (1947). Extension of Pearson's probability distributions to two variables, I, II, III, IV. In *Nederl. Akad. Wetensch. Proc., Ser. A*, volume 50, pages 1063–1070. 5, 17

Van Uven, M. (1948). Extension of Pearson's probability distributions to two variables, III. *Proc. Roy. Acad. Sci. Amsterdam*, 51:41–52, 191–196. 5, 17

von Neumann, J. and Morgenstern, O. (1953). *Theory of Games and Economic Behavior*. Princeton University Press. 100

Webster, M., Forest, C., Reilly, J., Babiker, M., Kicklighter, D., Mayer, M., Prinn, R., Sarofim, M., Sokolov, A., Stone, P., et al. (2003). Uncertainty Analysis of Climate Change and Policy Response. *Climatic Change*, 61(3):295–320. 73

Weinstein, N. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5):806–820. 98

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83. 63

Willink, R. (2008). A closed-form expression for the Pearson Type IV distribution function. *Australian & New Zealand Journal of Statistics*, 50(2):199–205. 2, 33, 89

Winiwarter, W. and Rypdal, K. (2001). Assessing the uncertainty associated with national greenhouse gas emission inventories: a case study for Austria. *Atmospheric Environment*, 35(32):5425–5440. 73

Womack, K. (1996). Do Brokerage Analysts' Recommendations Have Investment Value? *The Journal of Finance*, 51(1):137–167. 105

Yaari, M. (1987). The Dual Theory of Choice under Risk. *Econometrica*, 55(1):95–115. 101

Yuan, C. and Mao, X. (2004). Convergence of the Euler–Maruyama method for stochastic differential equations with Markovian switching. *Mathematics and Computers in Simulation*, 64(2):223–235. 130

Zivaripiran, H. (2005). *DDVERK90: A User-Friendly Implementation of an Effective DDE Solver*. PhD thesis, University of Toronto. 114

Zolotarev, V. (1986). *One-dimensional stable distributions*. American Mathematical Society. 90