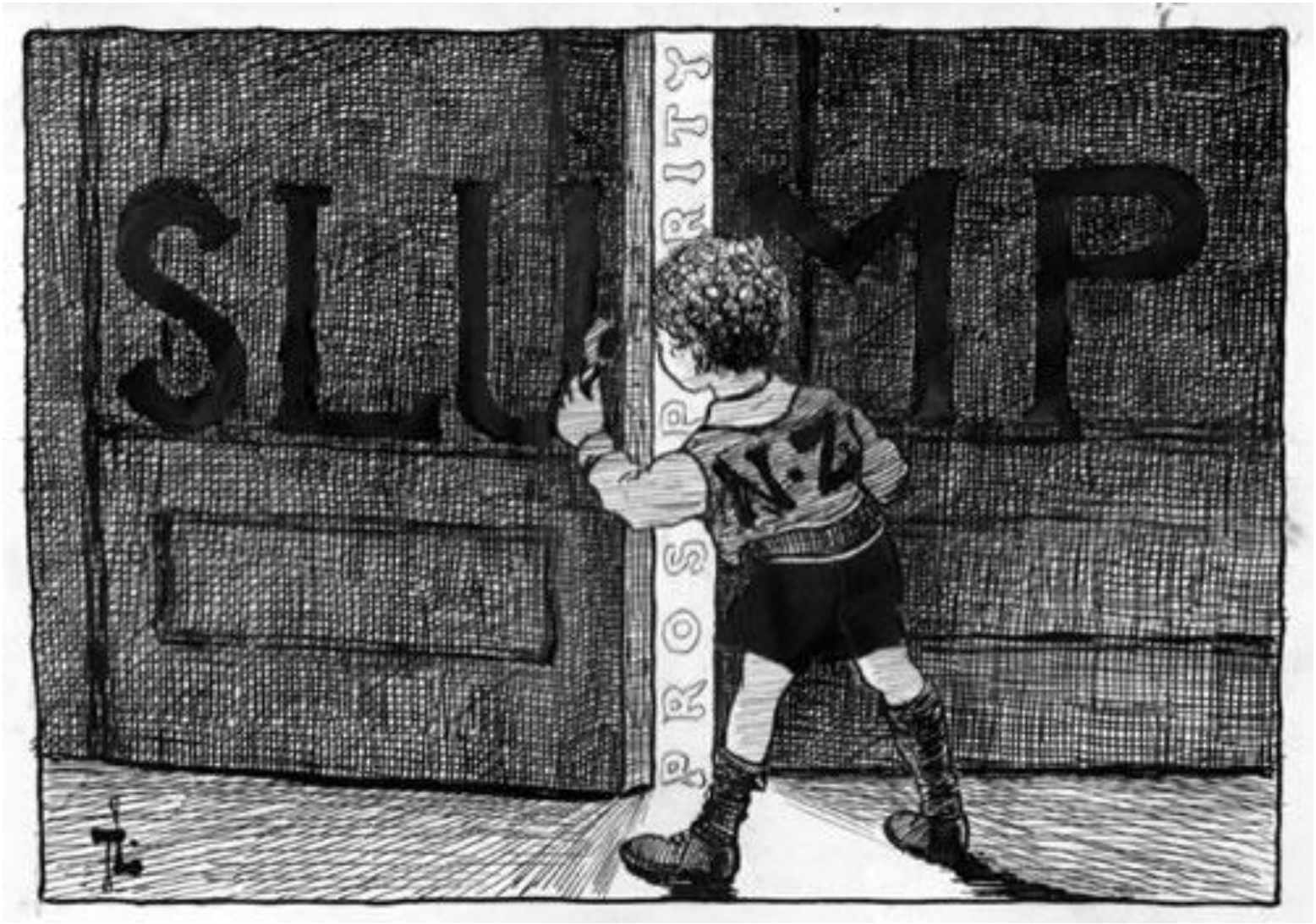
Using a copula-based model of GST data to visualise the New Zealand economy

Richard Vale

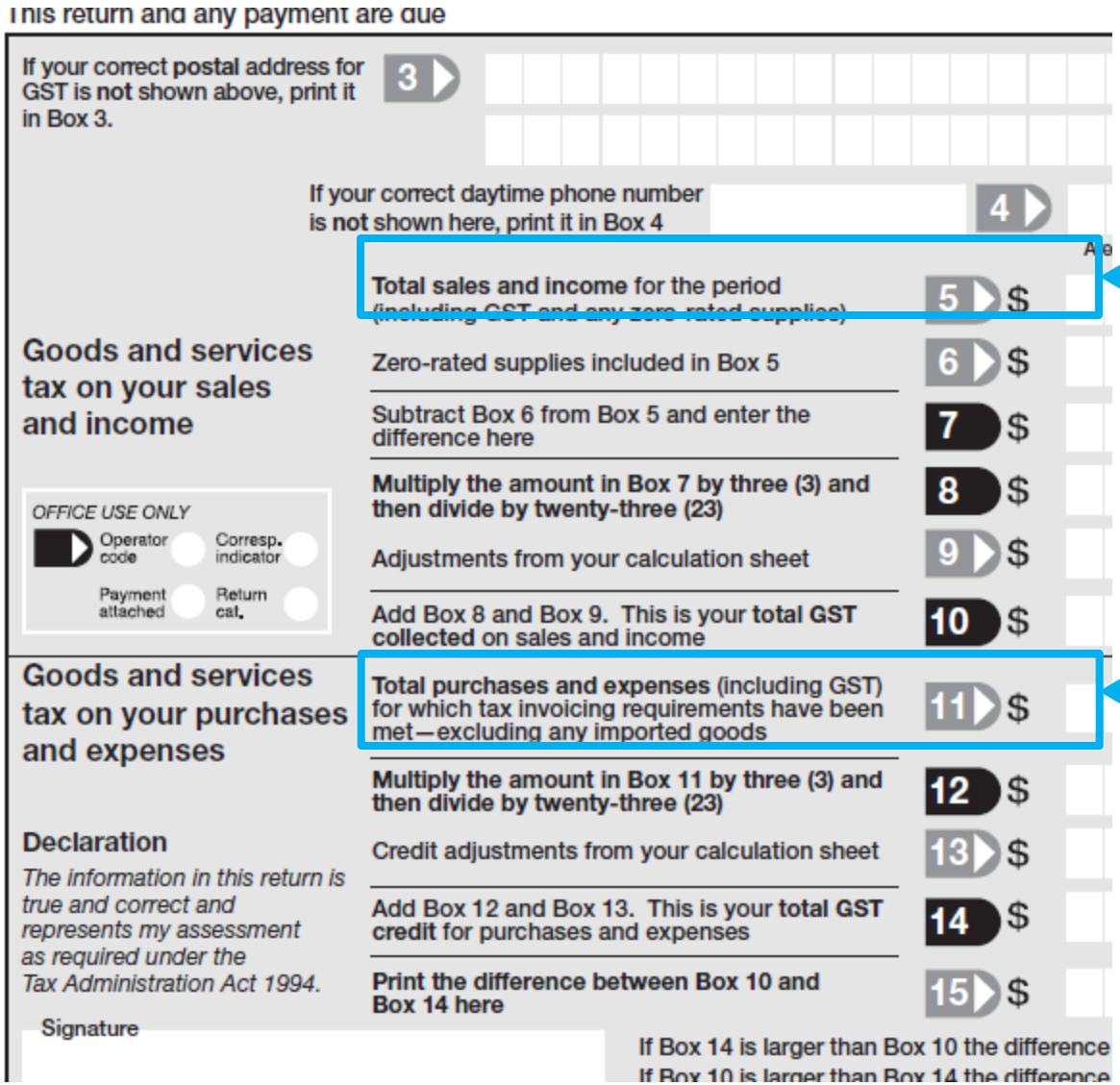
Inland Revenue
Te Tari Taake

Trevor Lloyd, 'A gleam of hope', *Zealandia*, 1933

''A gleam of hope' cartoon, 1933', URL: http://www.nzhistory.net.nz/media/photo/gleam-hope-cartoon-1933, (Ministry for Culture and Heritage), updated 20-Dec-2012



GST – a 15% tax on goods & services in NZ



This return and any payment are due

If your correct **postal** address for GST is not shown above, print it in Box 3. 3

If your correct daytime phone number is not shown here, print it in Box 4 4

Goods and services tax on your sales and income

Total sales and income for the period (including GST and any zero-rated supplies)	5	$
Zero-rated supplies included in Box 5	6	$
Subtract Box 6 from Box 5 and enter the difference here	7	$
Multiply the amount in Box 7 by three (3) and then divide by twenty-three (23)	8	$
Adjustments from your calculation sheet	9	$
Add Box 8 and Box 9. This is your **total GST collected** on sales and income	10	$

OFFICE USE ONLY
Operator code · Corresp. indicator · Payment attached · Return cat.

Goods and services tax on your purchases and expenses

Total purchases and expenses (including GST) for which tax invoicing requirements have been met—excluding any imported goods	11	$
Multiply the amount in Box 11 by three (3) and then divide by twenty-three (23)	12	$
Credit adjustments from your calculation sheet	13	$
Add Box 12 and Box 13. This is your **total GST credit** for purchases and expenses	14	$
Print the difference between Box 10 and Box 14 here	15	$

Declaration

The information in this return is true and correct and represents my assessment as required under the Tax Administration Act 1994.

Signature

If Box 14 is larger than Box 10 the difference
If Box 10 is larger than Box 14 the difference

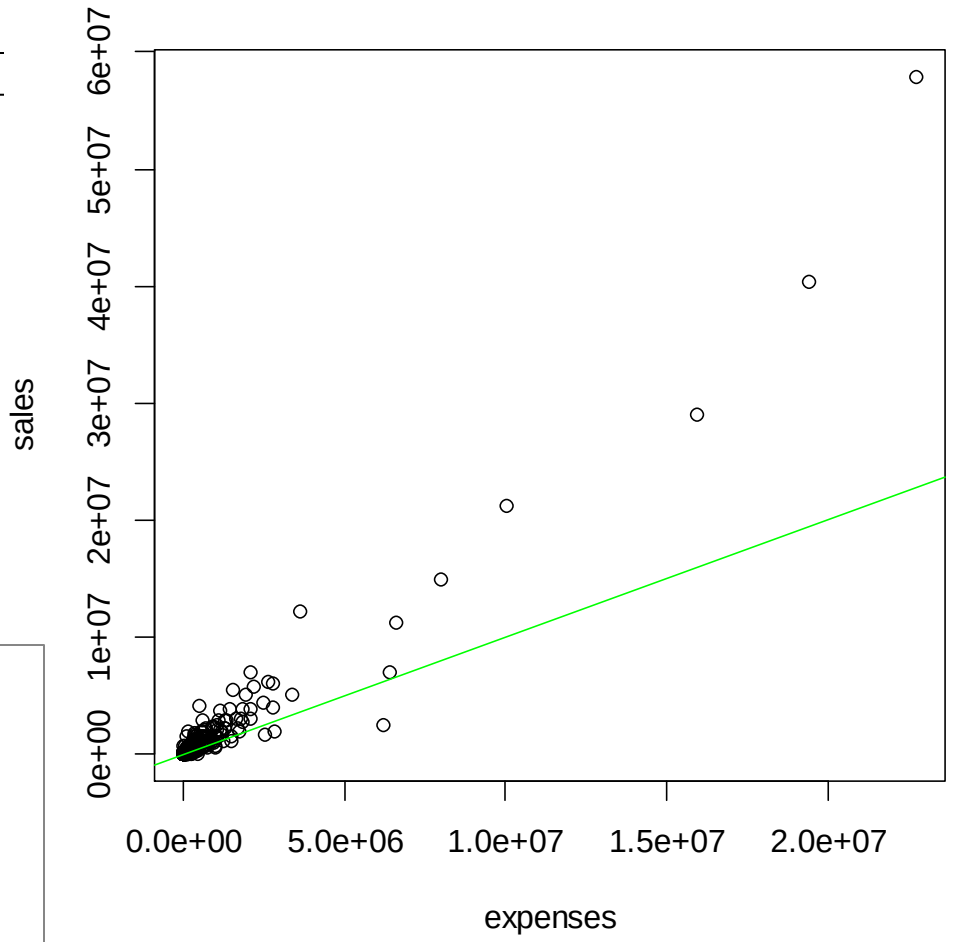
"sales"

Every business has to record its sales (income) and expenses on the GST form

"expenses"

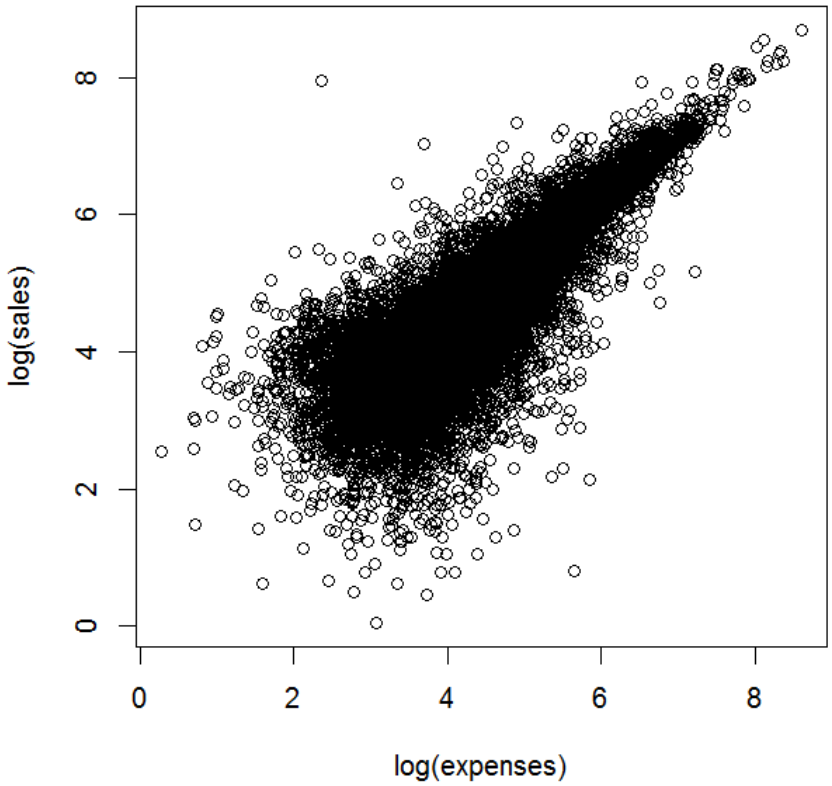
Data:

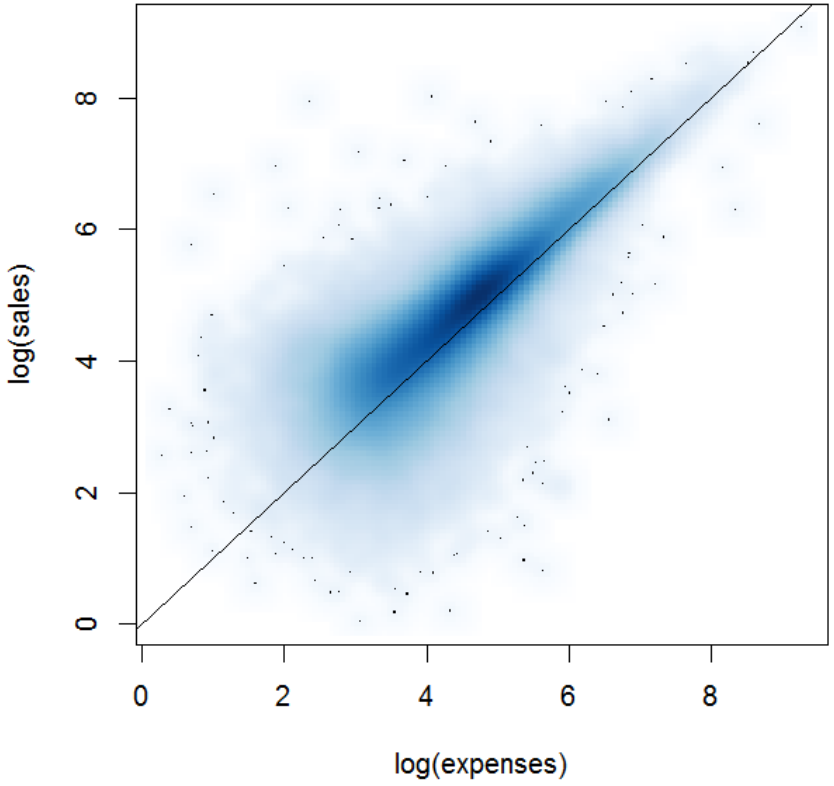
Name	expenses	sales
Small Coffee Co	4323	15357
United Poultry	23388	41494
Xmas Holdings	30450	0
OzCorp	5426038	1782330
…		



Plot of raw data with the line $y = x$. Most businesses (about 80%) have sales > expenses.

`smoothScatter()`





Aim:

- Reduce the dimension by fitting a model with a few parameters.
- Use the time series of the model parameters to visualise how the joint distribution changes from month to month

One way of fitting a bivariate distribution is to use a **copula**. A copula is just a function of the form

$$\int_0^x \int_0^y c(s,t)ds\ dt$$

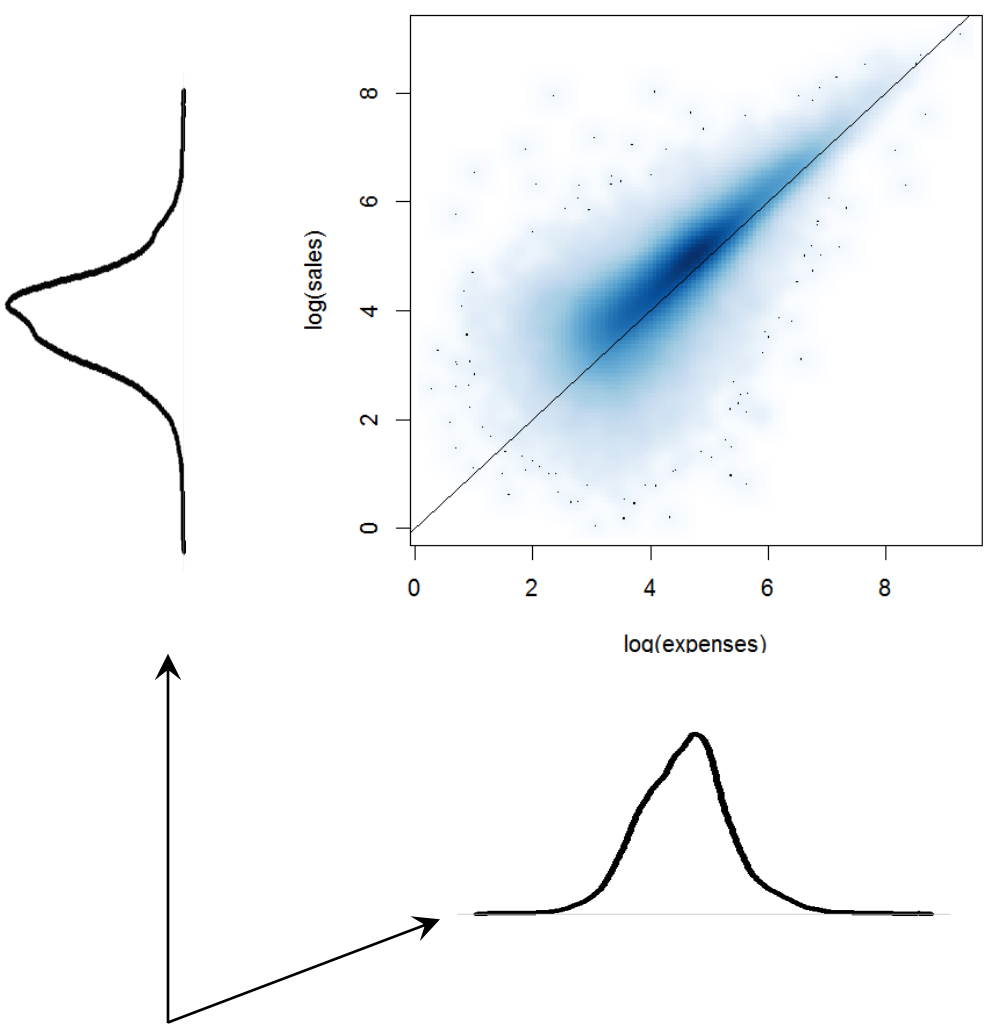
Where $c(s,t)$ is a density on the unit square with uniform marginals.

Sklar's Theorem: every bivariate distribution has the form

$$P(X \leq x, Y \leq y) = C(F_X(x), F_Y(y))$$

where C is a copula. So you can fit the marginals and copula separately!

(this may be statistically questionable)



Treat marginals as lognormal (4 parameters)

Fit (rotated) **Joe Copula** with lognormal marginals. Given by

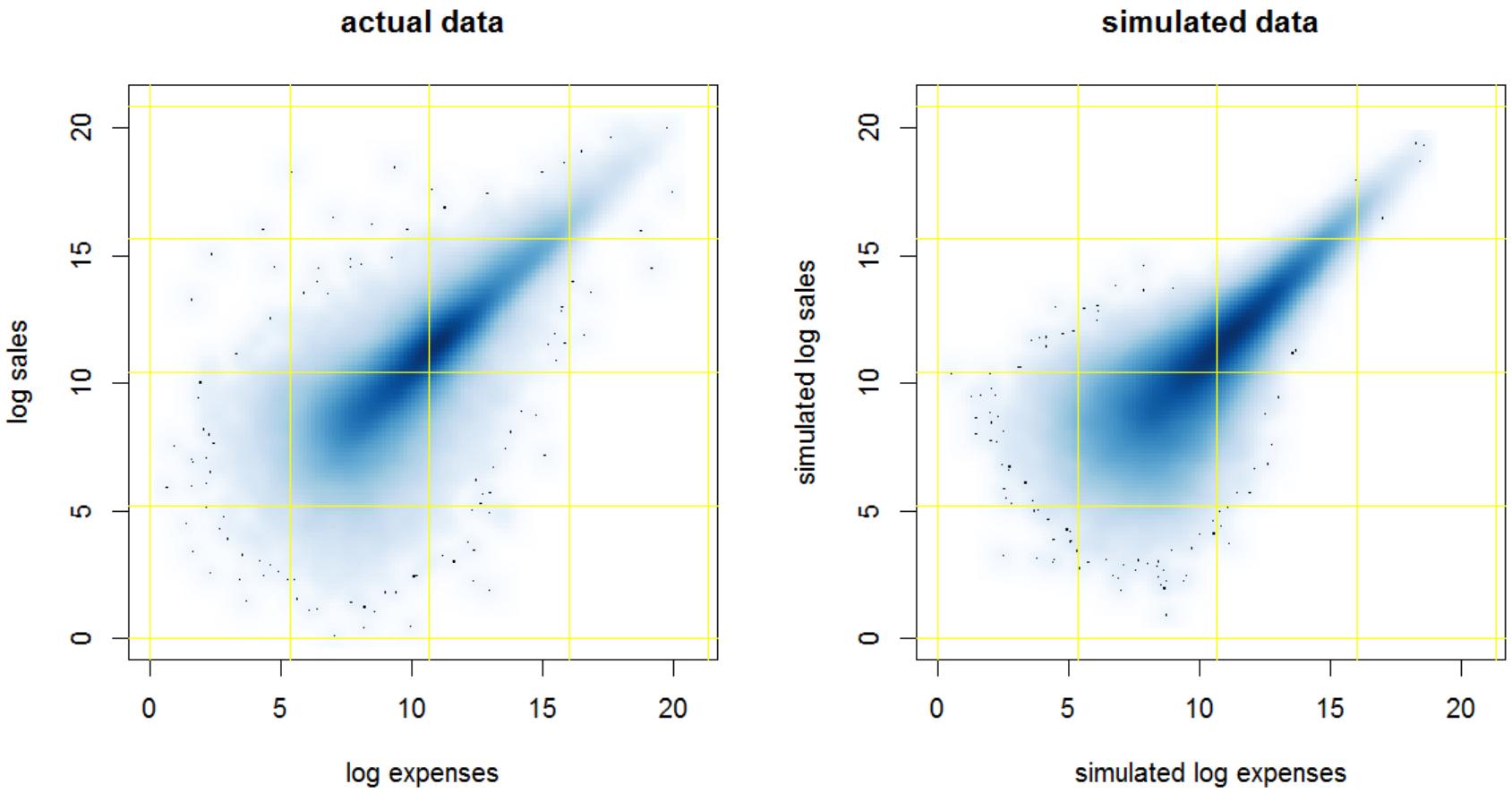
$$C(x, y) = \varphi^{-1}(\varphi(x) + \varphi(y))$$

where

$$\varphi(x) = -\log(1 - (1 - t)^{\theta})$$

(1 parameter) using `VineCopula` package.

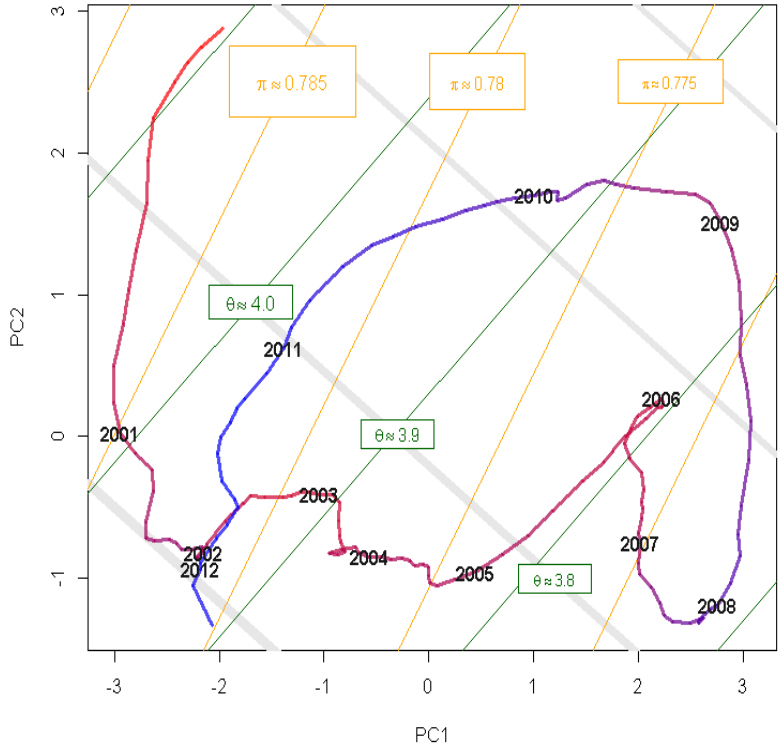
Finally, include a parameter π for the proportion of filers with sales > expenses (this is necessary because copulas fitted by software are symmetric about the line $y = x$).

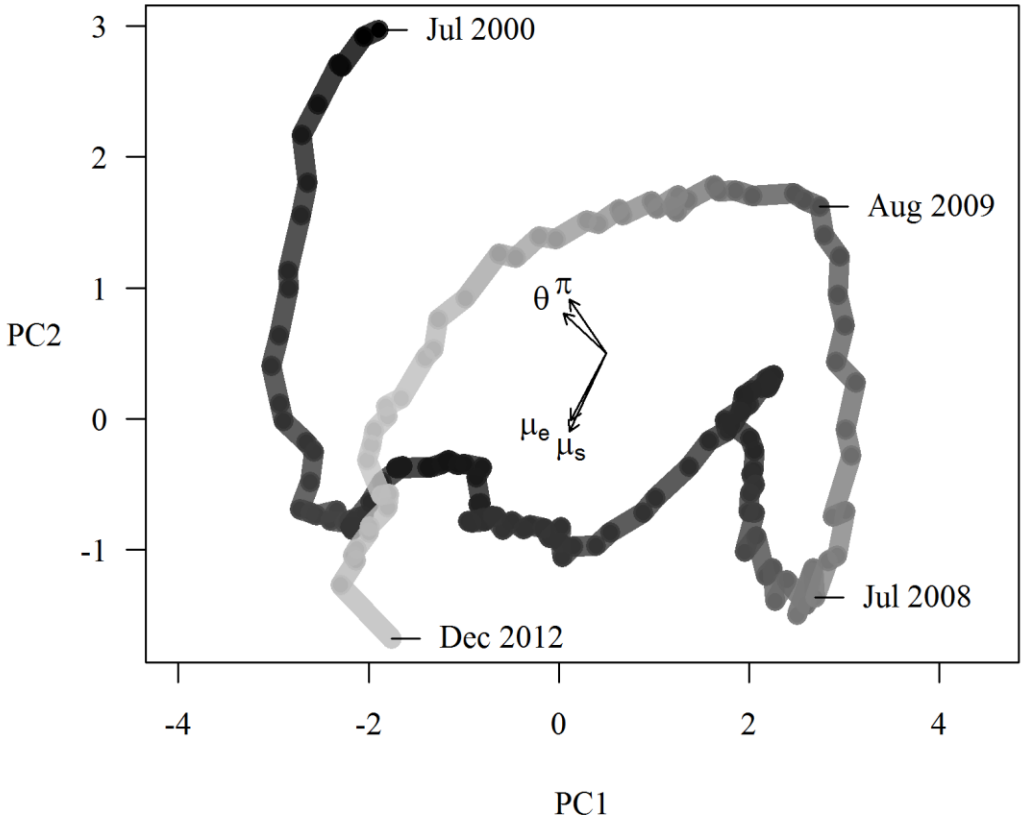


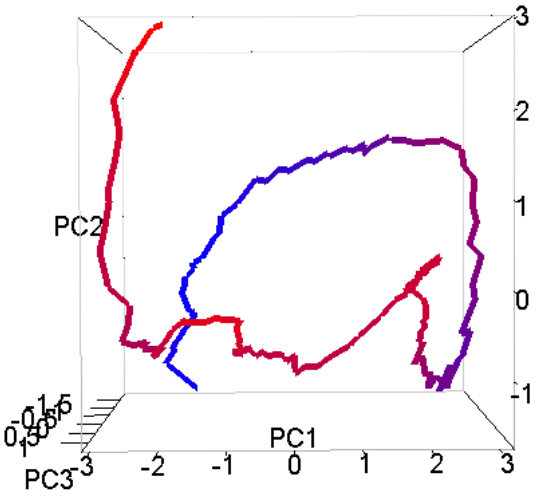
(Measures of goodness-of-fit are not trivial. See:

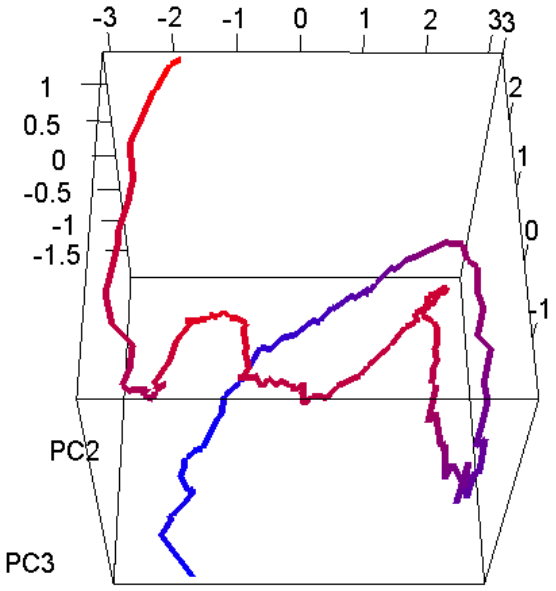
C. Genest, L.-P. Rivest, 1993. Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.* 88, 10341043.)

- Fit to each month from Jan 2000 to Jul 2013.
- Adjust for inflation.
- Take moving average (because of seasonality).
- Use PCA to reduce from six to three dimensions.
- Plot the first 2 principal components (using colour for the third dimension).

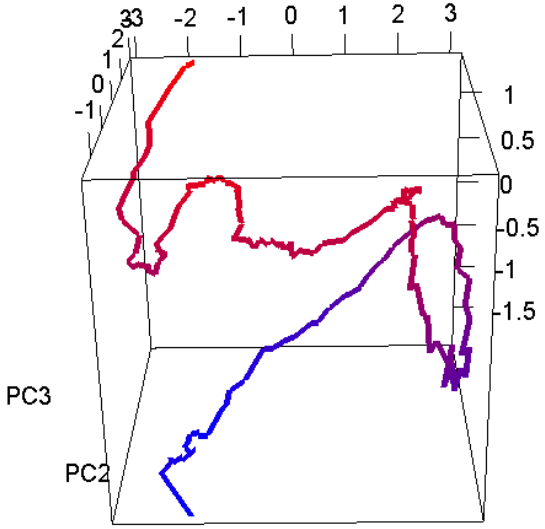


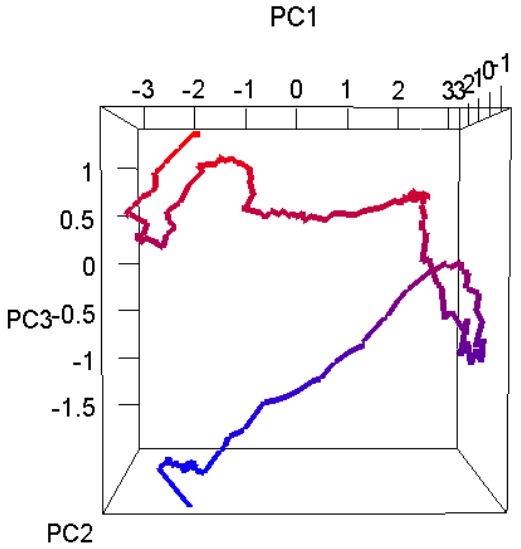




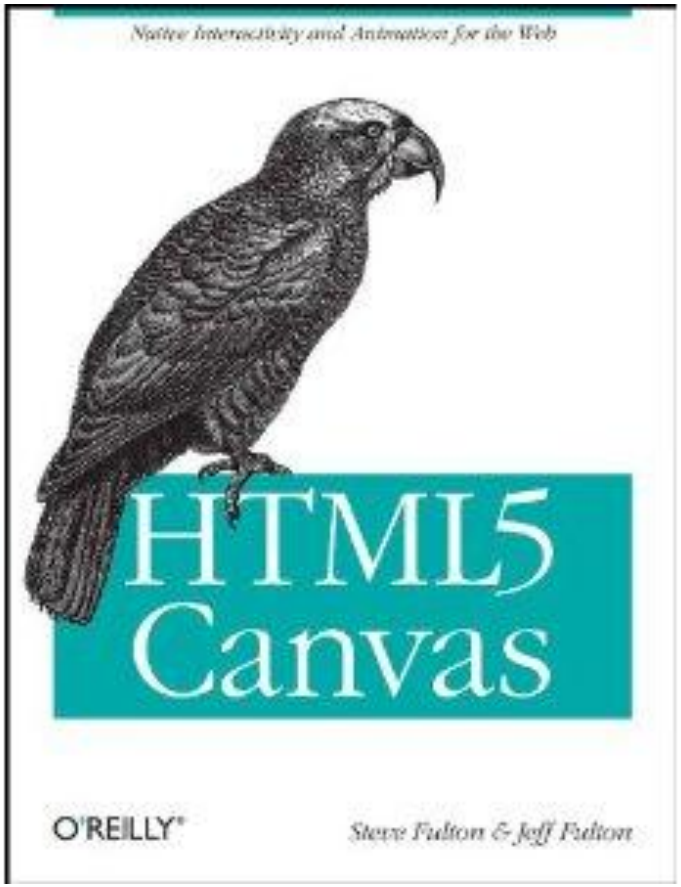


Visualisation using `rgl` package.
More of a spiral than a “cycle”.





Interactive version in html/JavaScript using the `canvas` element (quite like R graphics)



html:

```
<canvas id="canvas1" width="700",
height="620“>
</canvas>
```

…

JavaScript:

```
context = canvas1.getContext("2d");
context.fillStyle = "white";
context.fillRect(480, 12, 215, 80);

context.beginPath();
context.moveTo(5,5);
context.lineTo(5, ts.height-5);
context.stroke();
```

- Very small file size (e.g. compared to R `animation` package).
- Single file rather than a collection of .png files.
- You can make games with it.
- Only works in newer browsers (not IE).
- Entire screen must be redrawn on every frame (unlike Flash).
- Alternatives exist (e.g. svg; used by the popular d3.js library)

GST Animation - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Search Bookmarks and History

Google

Most Visited Getting Started Latest Headlines

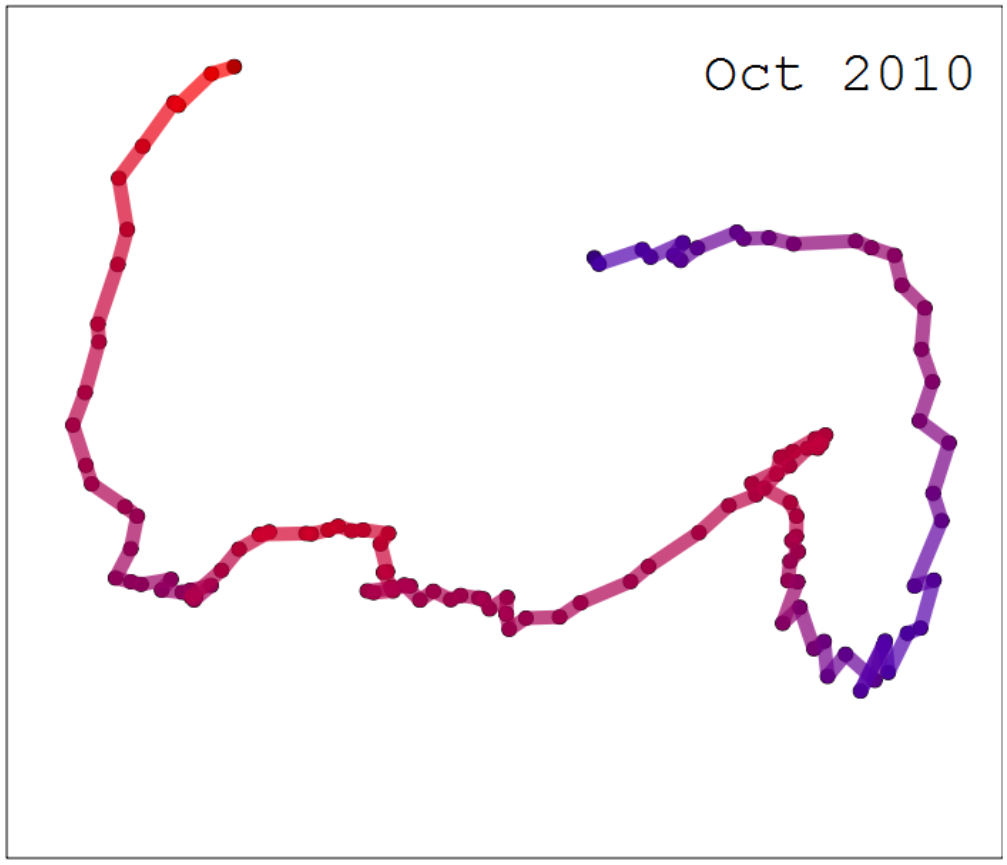
GST Animation

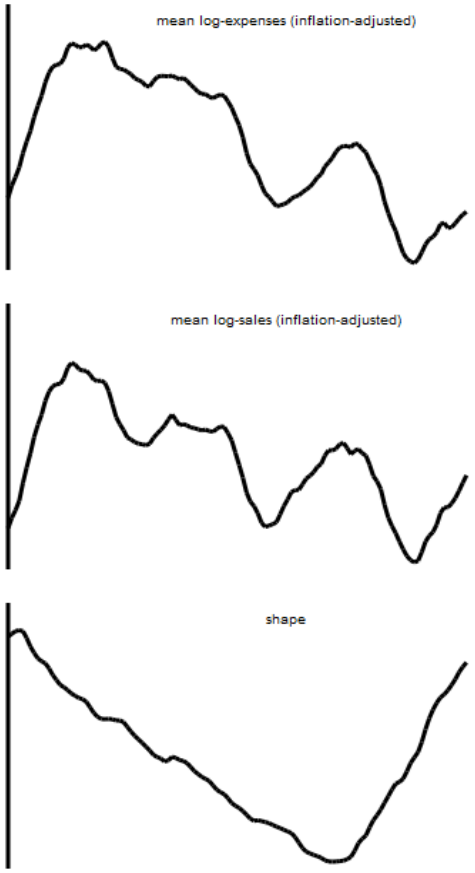
Evolution of GST expenses and sales model, 2001-2012

HTML5 animation by Richard Vale, Modelling. This page uses the <canvas> element, which is not available in older browsers. If you cannot view the animation, try opening the page in the latest version of Firefox.

Click for further information

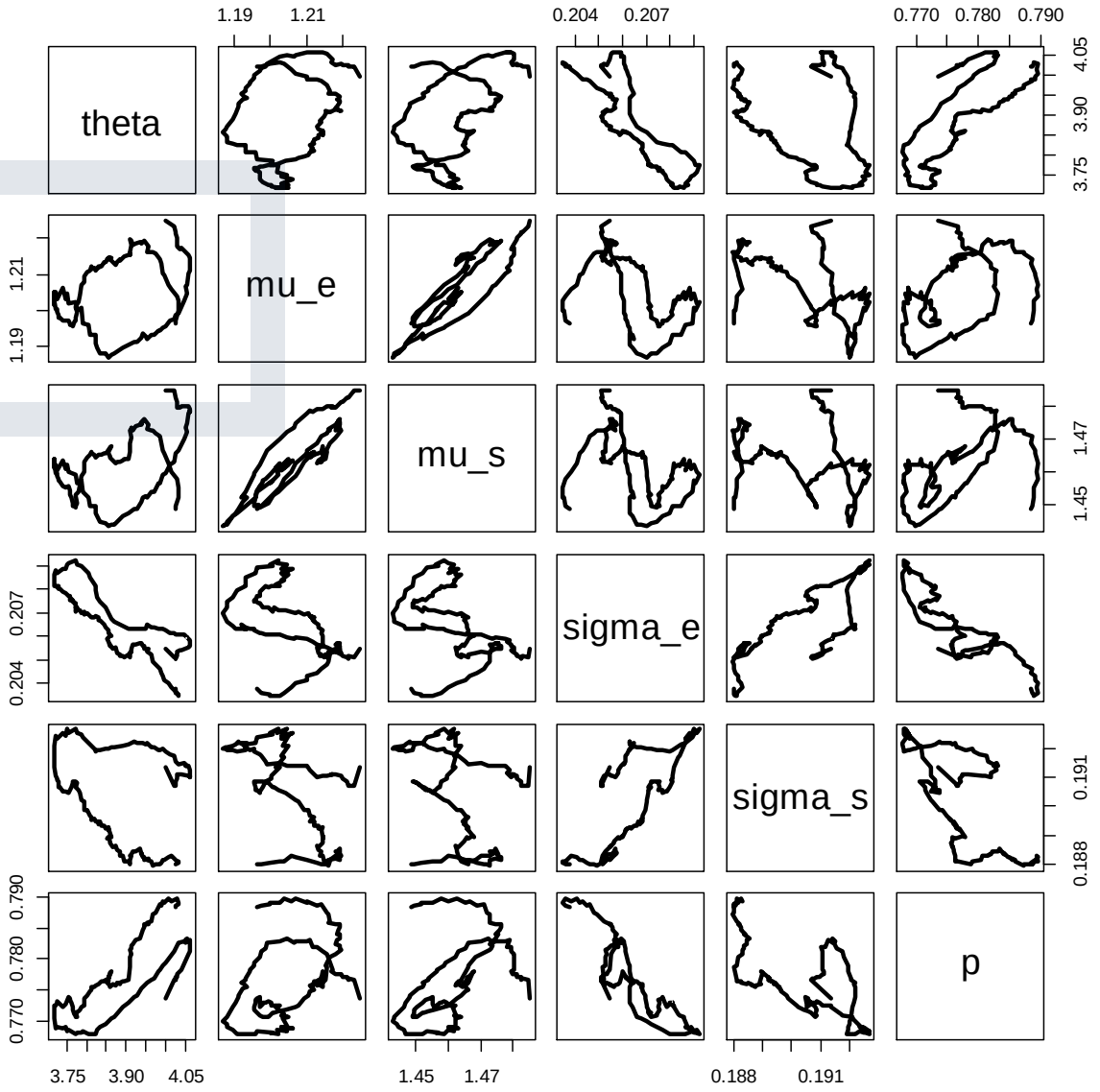
Animation controls: slower faster pause start reset reverse time

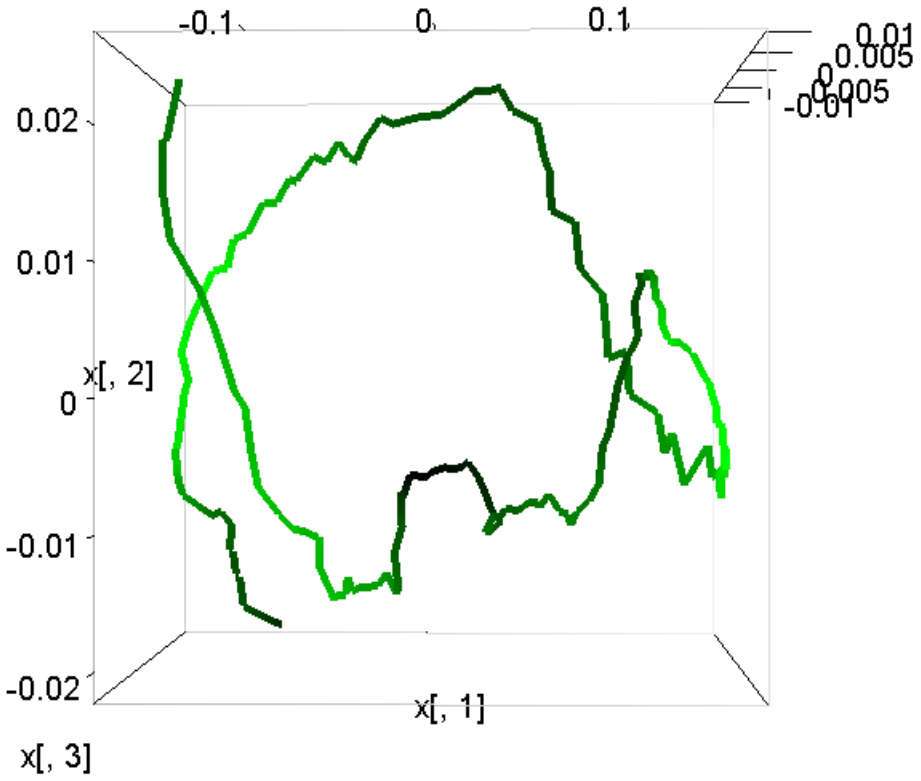


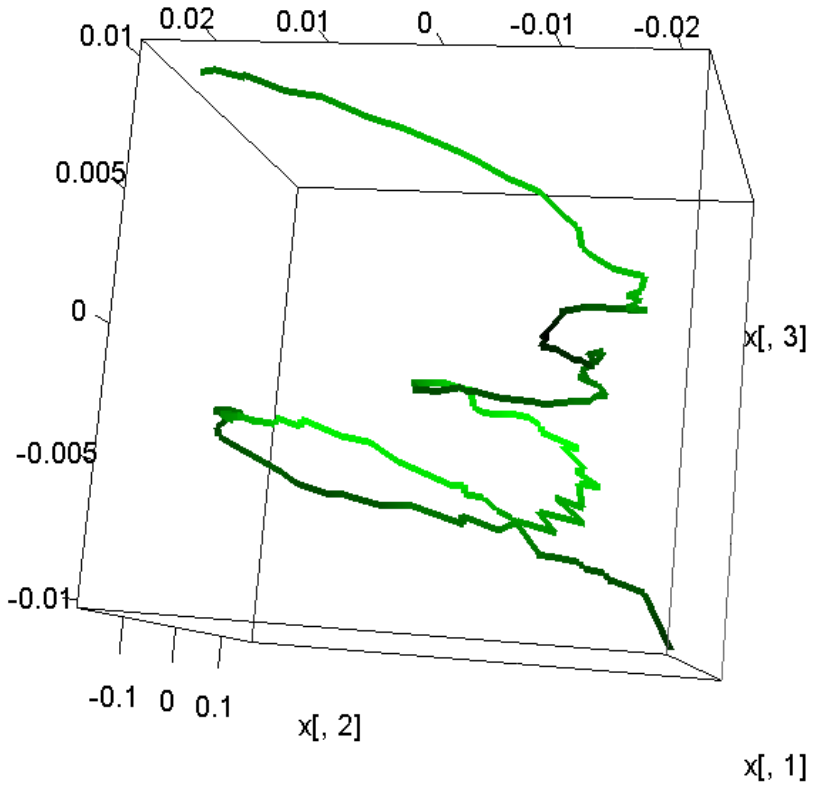


Done

Appendices


theta
mu_e
mu_s
sigma_e
sigma_s
p
1.19
1.21
0.204
0.207
0.770
0.780
0.790
3.75
3.90
4.05
1.45
1.47
0.188
0.191

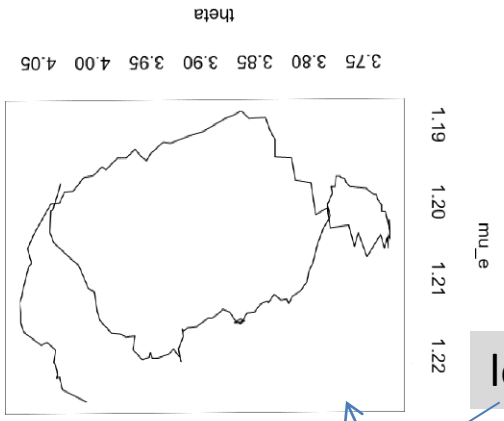




PCA on covariance matrix visualised using `rgl`

$PC1 \approx -\theta$
$PC2 \approx -\mu_e - \mu_s$
$PC3 \approx \pi$
$PC4 \approx \mu_s - \mu_e$

} Much the same as the correlation matrix version



looks much the same shape as this plot of theta vs mu_e