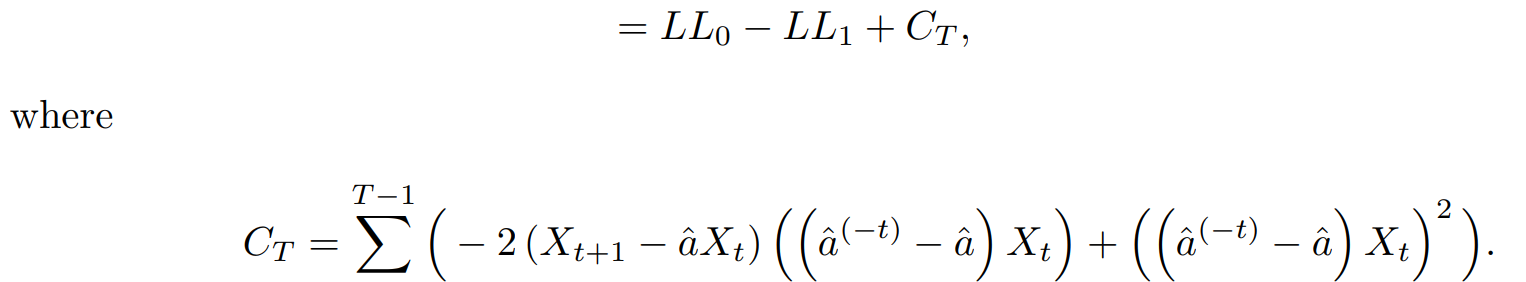
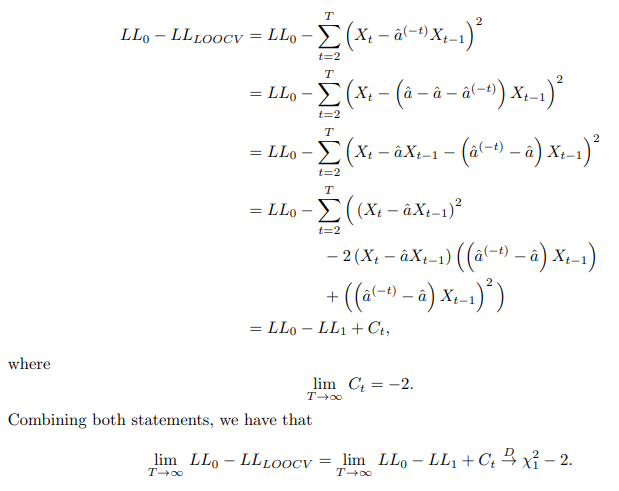
Prep Meeting Week 33

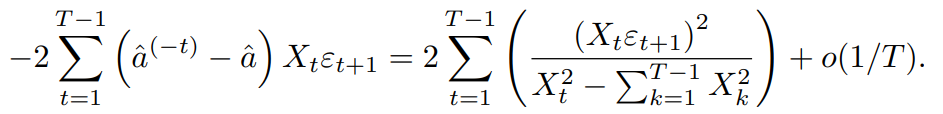
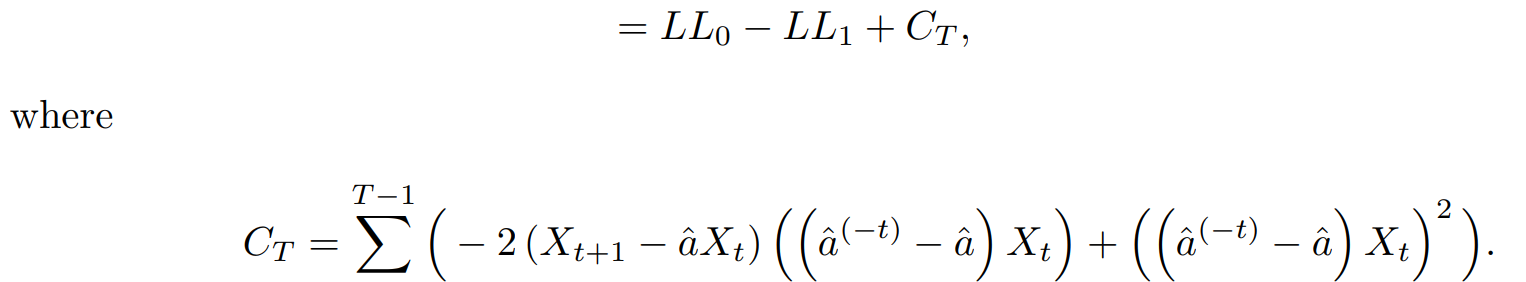
# Further Investigating the difference

Have rewritten our difference to

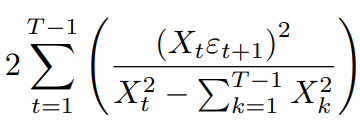




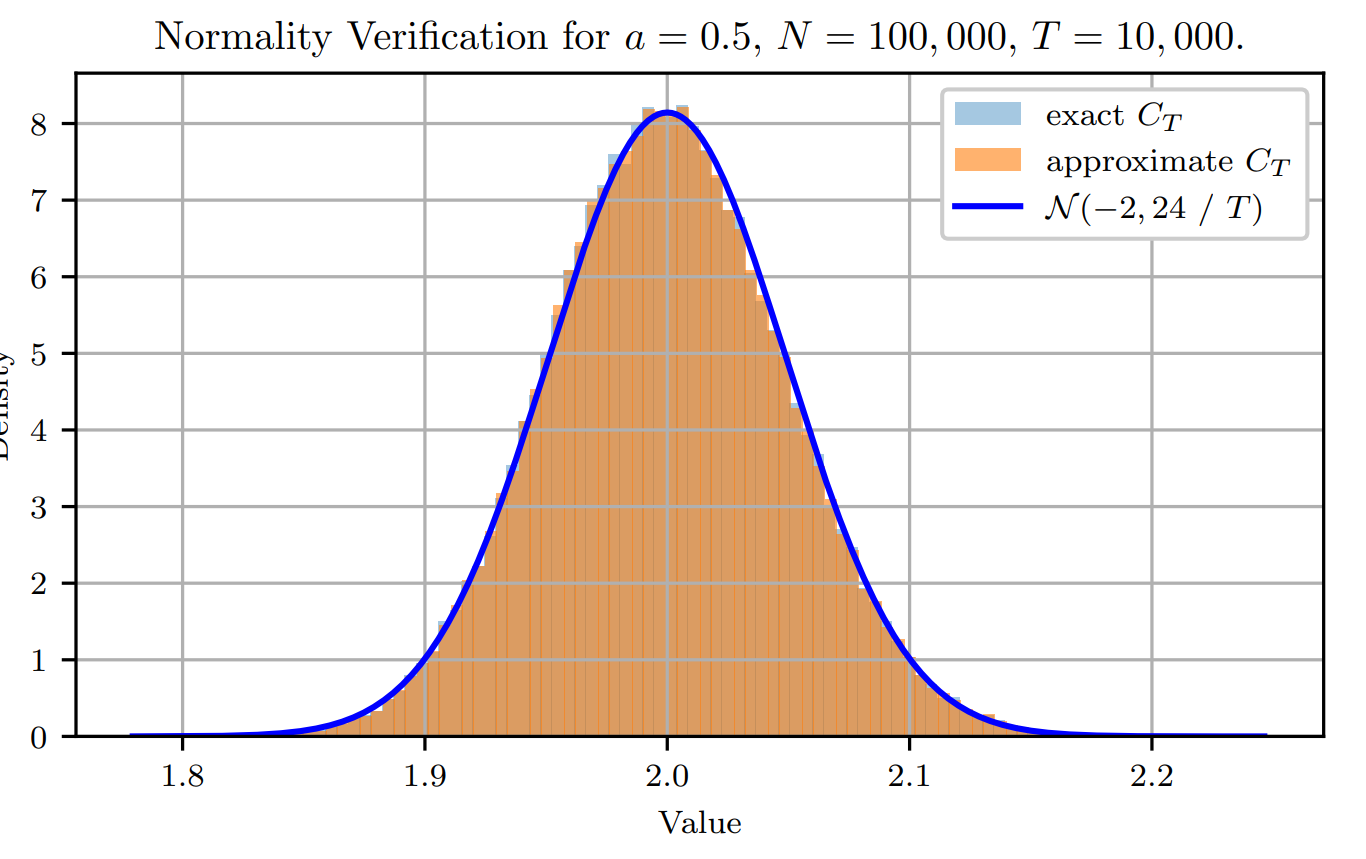
Now, we know that LL\_0 – LL\_1 converges to a chi squared distribution with one degree of freedom. Now, we know that *CT* is equal to



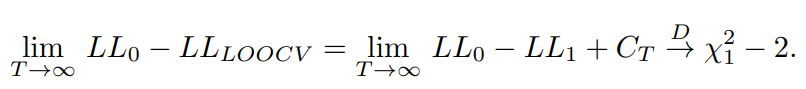
So, as *T* approaches infinity, the only contributing component will be

.

To verify this, we have plotted the histograms of the exact *CT* and the leading components of *CT* for values *T = 10,000*, where we have *N = 100,000* samples.



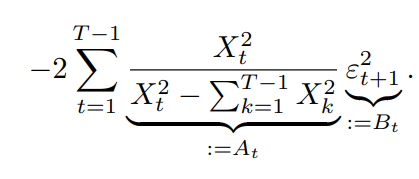
We see that the exact and approximate *CT* overlap almost perfectly. Moreover, the distribution seems to be a normal distribution with a mean of *-2* and more importantly, a variance of *24 / T*. Therefore, we can conclude that as *T* tends to infinity, *CT* tends to -2. And therefore,



Final part of derivation is quite difficult, and also the argument that the variance is of the order 1 / T is “empirical”, preferably, we would like to derive the approximate *CT* further, but difficult to say how.

**Attempts at further derivation.**

Tried to decompose the sum into two parts,



Where A\_t and B\_t seem independent, in the limit at least. Therefore, we saw that the mean of A\_t was 1 / T, and the mean of B\_t was 1. Using linearity of expectation, we got that the mean was -2 \* T / T = -2, as expected.

For the variance, things were more difficult, the variances separately allowed us to conclude that the variance of A\_t B\_t was 8 / T^2. Now, summing these T variables would yield a variance of 8 / T, which is **not** the 6 / T that we expected. This is due to the correlation between A\_tB\_t and A\_t’B\_t’, but accounting this this is rather difficult.

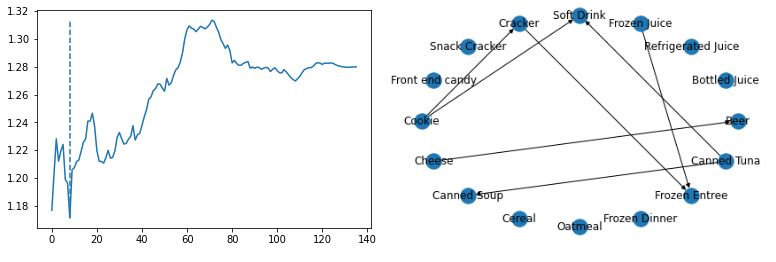
Question is how to derive this more neatly, rather than just stating this normality at the end?

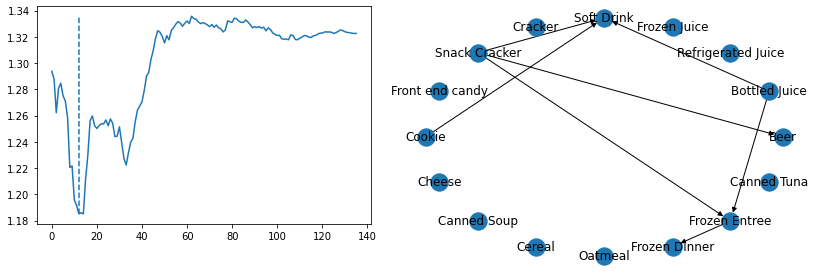
# Dominick’s Finest Foods Data

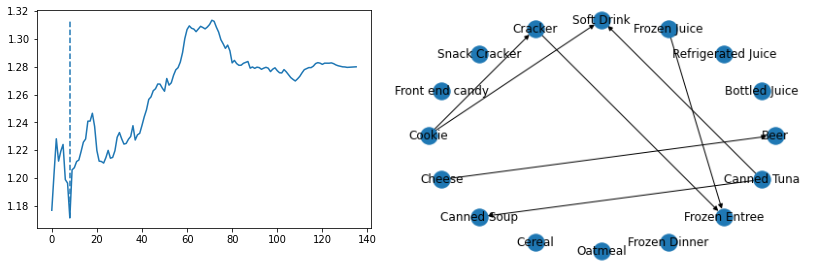
Investigated Dominick’s Finest Foods data, which is a dataset suitable for VAR(1) modelling and appeared in some papers, seemed to be an interesting real-life dataset where some modelling assumptions might not be met, interesting to see how our methods work.

From the experiments of Gelper et al., they derived this directed network.

Based on our analysis, we derive this directed network for store **8,** **21, 68**.

****

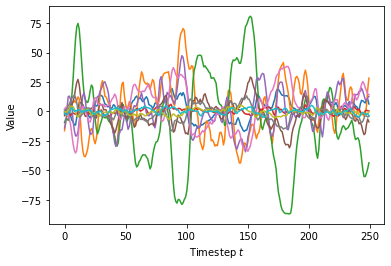
****

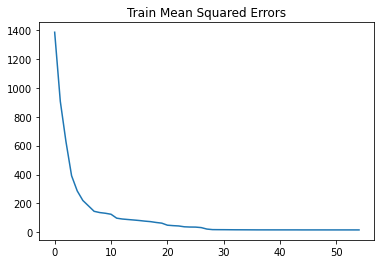
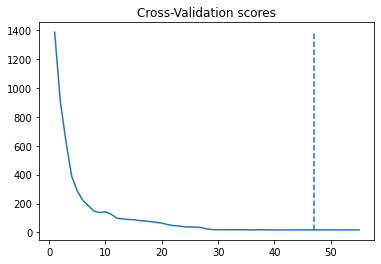
****

# LOOCV Performance on mismatch

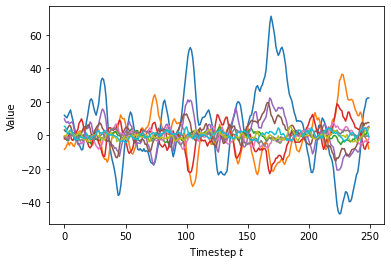
Performance seems okay for increasing noise variance.

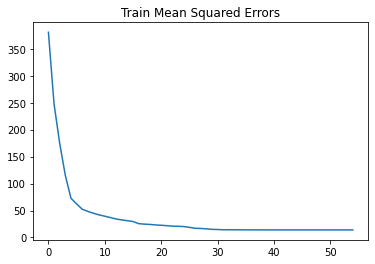
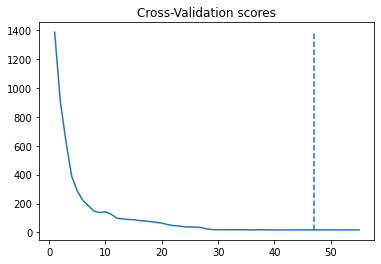
Performance seems not okay for when the noise is a VAR(1) model itself, although the model selection does not seem to be very difficult.



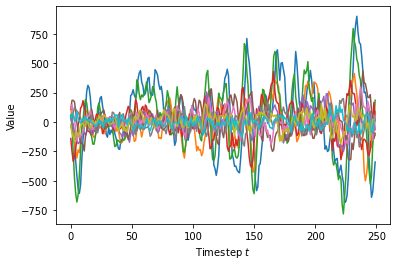


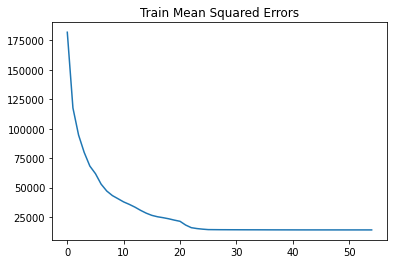
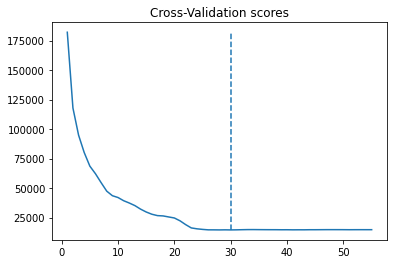
**Now with no a diagonal coefficient matrix for generating noise, so *p* separate AR(1) processes**

****

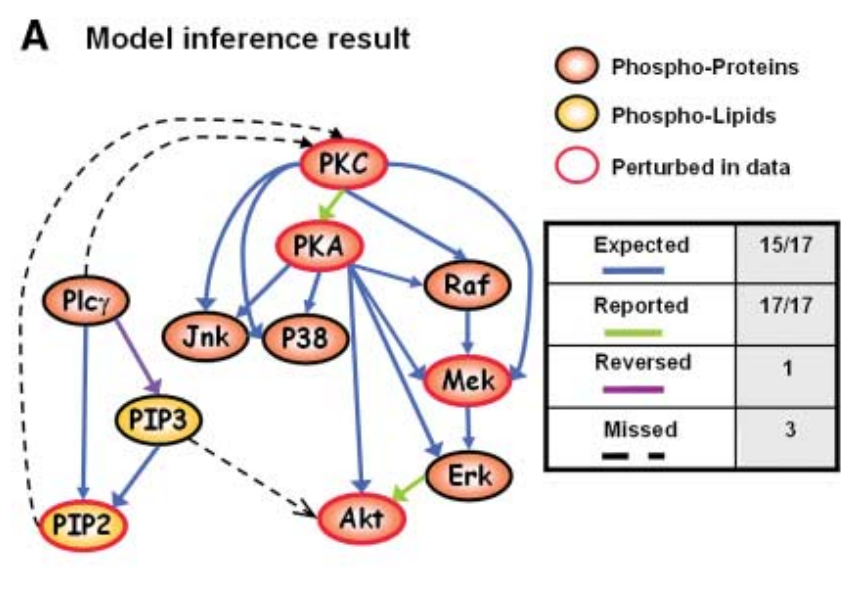
****

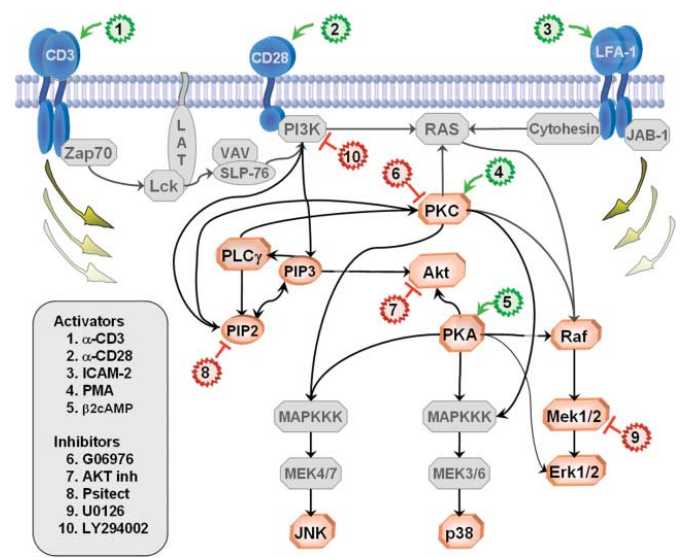
**Using a steadily increasing noise level, but still “i.i.d” Gaussian zero mean variables**

****

****

# Sachs Dataset Results





**Inferred Results using OMP, NO TEARS, FGES, and other methods of ours.**