INFORMATION STRUCTURES

Prepared for the 2000 Investment Conference
Andrew Smith
of the
Applications of Financial Theory Working Party

andrew.d.smith@bacon-woodrow.com

Abstract:

This paper briefly investigates some ways of modelling information structures in capital markets. We introduce the new concept of white thunder, and apply it to equity markets.

Part I: Orientation

A Common Modelling Problem

You are building an asset-liability model for an insurance company. You know that next year's expenses will be about £100,000, and you also expect them to vary according to inflation over the next year. You use an off-the-shelf actuarial model to generate simulated macro-economic outcomes.

To your surprise, you find that, in one simulation in 4, the expense projection is above £105,000 or below £95,000. Even more extreme events can happen; we fall outside the range £90,000 to £100,000 about one simulation in 50.

This seems very strange. These expenses have become one of the most variable items in the projection. And yet the risks supposedly only reflect the risk of inflation in the underlying economy. Even the government manages to forecast inflation a year ahead within 1% or so. So why does the model seem so hopeless?

So you go back to check the model, and examine historic inflation data. To your surprise, you find that the model fits the data pretty well. There seems to be little scope for improving the fit from a statistical perspective. Certainly, there is nothing in past inflation data to suggest that future inflation could be predictable to with 1% a year ahead. So we have the puzzle of a model which passes all the statistical tests, and yet the insurance company is unhappy with the model you have delivered.

We will see that the problem here is not one of statistics, but one of information. Much of the information regarding next year's inflation is already in the market. However, under the time series inflation model, that information appears in next year's residual term. What we need is some way of allowing future error terms to be predicted, without undermining their independence from each other.

A Tale of Two Exchanges

We illustrate this problem further with a tale of two stock exchanges. These are located in two towns: Donner and Blitz.

Both these towns publish a daily share price index. In each case, to the statistical eye, the log index seems to perform a standard Gaussian random walk. The observant also notice that these indices are exactly the same - except that the Donner index always lags the Blitz index by a day. So to find out the Donner index on a particular day, it suffices to look up the Blitz index from the previous day.

Now, given the speed of modern communications, we might be rather surprised at the behaviour of the Donner market. We might wonder why arbitrage traders did not pile in and out of the market each day until the lag against the Blitz market disappeared. However, we would not suspect anything was odd until we looked at Blitz. Donner, viewed on its own, appears an ordinary stock market with ordinary random walk behaviour. Statistical analysis of prices would throw up nothing - absolutely nothing - to give us a clue.

Separating Stationary Laws from Information Structures

In both these examples, we need to make a distinction between stationary laws and information structures. The stationary law of a stochastic process is the probability law driving that process. It is an unconditional probability law, and if the process is stationary then we can hope to estimate the probability law by statistical analysis of past history.

All this statistical analysis does not rule out the possibility that someone might be able to make better forecasts, by allowing for other information. But, to determine if this is so, we must look outside the initial series to other sources of information. We now consider some practical models to help us achieve this.

Part II: Mathematical Models of Information

Information Structures and Sigma-Algebras

In probability theory textbooks, information structures are captured by a structure known as a "sigma algebra". Without getting into measure theoretic technicalities, we note that for most financial models, the relevant sigma algebra is an infinite set of (mostly) infinite sets. This is not the kind of structure that readily lends itself to empirical observation. However, the fact that mathematicians see the need to introduce these complexities indicates that information structures may be important, in any area where probabilities are used. In this note we seek to construct a more concrete approach to information structures.

Canonical White Thunder

Having described the idea of separating information and processes, we now describe our canonical white thunder process.

We start with a white noise process e_t, defined for all positive and negative integers t. We suppose these are a series of independent N(0,1) random variables. At time t, we know the values of e_t and all previous e's, but not the value of any future e's.

Let us pick a constant A, with $|A| < 1$. We define another process x_t inductively, as follows:

$$x_t = Ax_{t-1} + Ae_t - e_{t-1}$$

starting the induction at a very early value of t. By induction, we can show that

$$x_{t+h} = A^{h-1}(Ax_t - e_t) + Ae_{t+h} - (1-A^2)\sum_{j=1}^{h-1} A^{h-1-j} e_{t+j}$$

or, equivalently,

$$x_t = A^{h-1}(Ax_{t-h} - e_{t-h}) + Ae_t - (1-A^2)\sum_{j=1}^{h-1} A^{j-1} e_{t-j}$$

Letting the starting point tend back to minus infinity, we have

$$x_t = Ae_t - (1-A^2)\sum_{k=1}^{\infty} A^{k-1} e_{t-k}$$

The process x then defines *canonical white thunder*.

Probability Law for White Thunder

What is the probability behaviour of our white thunder process? With a little effort, we can demonstrate that the x_t are also independent N(0,1) random variables. For those who like that sort of thing, here's the proof.

We can calculate the variance as follows:

$$\begin{aligned}\mathbf{Var}(x_t) &= \mathbf{Var}\left[Ae_t - (1-A^2)\sum_{k=1}^{\infty} A^{k-1}e_{t-k}\right] \\ &= A^2 + (1-A^2)^2\sum_{k=1}^{\infty} A^{2(k-1)} \\ &= A^2 + (1-A^2) \\ &= 1\end{aligned}$$

and if $h \geq 1$, we have

$$\begin{aligned}\mathbf{Cov}(x_t, x_{t+h}) &= \mathbf{Cov}\left[Ae_t - (1-A^2)\sum_{k=1}^{\infty} A^{k-1}e_{t-k}, Ae_{t+h} - (1-A^2)\sum_{k=1}^{\infty} A^{k-1}e_{t+h-k}\right] \\ &= \mathbf{Cov}\left[Ae_t - (1-A^2)\sum_{k=1}^{\infty} A^{k-1}e_{t-k}, -(1-A^2)\sum_{k=h}^{\infty} A^{k-1}e_{t+h-k}\right] \\ &= \mathbf{Cov}\left[Ae_t - (1-A^2)\sum_{k=1}^{\infty} A^{k-1}e_{t-k}, -(1-A^2)\sum_{k=0}^{\infty} A^{k+h-1}e_{t-k}\right] \\ &= -(1-A^2)A^h\left[1-(1-A^2)\sum_{k=1}^{\infty} A^{2(k-1)}\right] \\ &= 0\end{aligned}$$

This enables us to verify that the white thunder process is also a series of independent N(0,1) variables. No statistical test could possibly distinguish between the a true white noise process and our white thunder process.

The Embedded Autoregressive Process

Although both x_t and e_t are series of independent N(0,1) variables, sums of these series are not independent. Indeed, we can re-arrange the regression equation to give:

$$Ax_t - e_t = A[Ax_{t-1} - e_{t-1}] - (1-A^2)e_t$$

We can immediately recognise this as the law of a first order autoregressive process, for which the stationary distribution is:

$$\mathbf{Cov}(Ax_s - e_s, Ax_t - e_t) = (1-A^2)A^{|t-s|}$$

This is intriguing; although x_t is a series of independent N(0,1) variables we can express it as a sum of white noise and a mean reverting process. This aspect can lead to confusion – the fact that a mean reverting process enters somewhere in a model does not imply that the mean reversion will be evident from the final output.

Invertibility

We have shown how to construct a white thunder process from white noise. It is reasonable to ask whether the same idea can be applied in reverse. In other words, given observations only of the x_t, can we construct the e_t?

To do so would require recursive application of the equation relating e and x. On rearrangement, we have

$$e_t = \frac{e_{t-1} + x_t}{A} - x_{t-1}$$

Our problem here is that the recurrence relation is unstable – remember that $|A| < 1$. If we guess an initial e_t a very long way back, and then try to compute the e's recursively by substituting for the observed x's, the effect of any error in this initial guess grows exponentially as we move forward.

Thus, although we have two statistically indistinguishable processes, the information structures are different. We can deduce the x's from the e's, but not the other way round. In other words, there is more information in the e's than in the x's.

Prediction

The white thunder property of the process x_t only works because we can observe the original error series e_t as well.

As x_t is, on its own, simply a white noise process, there is no way that we could use the past of x_t to predict its future. But suppose we had the history of x and also the history of e – could we use this to forecast x?

The answer to this question is *yes*. We use $\mathbf{E}_t$ to denote the conditional expectation given x_t, e_t and all history of both processes prior to time t. Then, by substitution into the recurrence relation, we can show that for $h > 0$ we have

$$\mathbf{E}_t(x_{t+h}) = A^{h-1}\left[Ax_t - e_t\right]$$

It is this predictability which distinguishes white thunder from white noise.

PART III: CONSEQUENCES FOR ECONOMIC MODELS

Efficient Market Hypothesis

One of the most important hypothesis of modern finance is the efficient market hypothesis, or EMH. EMH states that market prices reflect available information, with varying definition of information. So any test of EMH must involve a specification of the information set to be considered.

It market prices fully reflect information, then price changes should be due to surprises, that is, new information. A major theme in the market efficiency literature involves testing whether market price changes are unpredictable (as implied by EMH) or whether they can be predicted in some way.

If, as is commonly found, market price changes have small correlations with changes in earlier periods, then this seems to support EMH. But there is a danger here. Even if prices can be shown to approximate a random walk, EMH does not follow. The price changes might be white thunder, in which case EMH would not hold. So to return to our original example, Blitz could have an efficient stock market, but Donner certainly does not.

Dividend Growth as White Thunder

It is commonly observed that real changes in dividends across different years do not seem to be highly correlated. So, at first sight, we might try to model changes in dividends as a white noise process.

If dividend changes were white noise, then future dividend changes would be independent of the past. If prices were to be computed by discounting dividends, and the discount rate were constant, we would then expect prices to be a constant multiple of dividends. In other words, the dividend yield should be constant.

In fact, of course, dividend yields fluctuate considerably over time, although many would argue that they should at least be a stationary process. The challenge is to reconcile this to the apparent random walk behaviour of dividends.

One explanation is that the changes in dividends are governed not by white noise but a mixture of white noise and white thunder. Specifically, let us suppose that x_t is a white thunder process, and that y_t is white noise, independent of x. Let us denote the real dividend index by D_t, and let us suppose that

$$\frac{D_t}{D_{t-1}} = \exp\left[\mu + \sigma_x x_t + \sigma_y y_t\right]$$

Or, alternatively, for $h \geq 0$

$$\log D_{t+h} = \log D_t + \mu t + \sigma_x \sum_{j=1}^{h} x_{t+j} + \sigma_y \sum_{j=1}^{h} y_{t+j}$$

Moments of the White Thunder Dividend Model

We now manipulate this to determine means and variances. We recall that:

$$x_{t+h} = A^{h-1}(Ax_t - e_t) + Ae_{t+h} - (1-A^2)\sum_{j=1}^{h-1} A^{h-1-j} e_{t+j}$$

so that sums of x satisfy:

$$\begin{aligned}
\sum_{j=1}^{h} x_{t+j} &= \sum_{j=1}^{h}\left[A^{j-1}(Ax_t - e_t) + Ae_{t+j} - (1-A^2)\sum_{i=1}^{j-1} A^{j-1-i} e_{t+i}\right] \\
&= \frac{1-A^h}{1-A}(Ax_t - e_t) + A\sum_{j=1}^{h} e_{t+j} - (1-A^2)\sum_{i=1}^{h-1}\sum_{j=i+1}^{h} A^{j-1-i} e_{t+i} \\
&= \frac{1-A^h}{1-A}(Ax_t - e_t) + A\sum_{j=1}^{h} e_{t+j} - (1-A^2)\sum_{i=1}^{h-1}\frac{1-A^{h-i}}{1-A} e_{t+i} \\
&= \frac{1-A^h}{1-A}(Ax_t - e_t) - \sum_{j=1}^{h}\left[1-(1+A)A^{h-j}\right] e_{t+j}
\end{aligned}$$

It follows that

$$\mathbf{E}_t\left[\sum_{j=1}^{h} x_{t+j}\right] = \frac{1-A^h}{1-A}(Ax_t - e_t)$$

and

$$\begin{aligned}
\mathbf{Var}_t\left[\sum_{j=1}^{h} x_{t+j}\right] &= \sum_{j=1}^{h}\left[1-(1+A)A^{h-j}\right]^2 \\
&= h - 2(1+A)\frac{1-A^h}{1-A} + (1+A)^2\frac{1-A^{2h}}{1-A^2} \\
&= h - \frac{1+A}{1-A}(1-A^h)^2
\end{aligned}$$

Substituting in, and using the normal moment generating function, we can deduce that

$$\mathbf{E}_t(D_{t+h}) = D_t \exp\left[\mu h + \sigma_x \frac{1-A^h}{1-A}(Ax_t - e_t) + \left(h - \frac{1+A}{1-A}(1-A^h)^2\right)\frac{\sigma_x^2}{2} + h\frac{\sigma_y^2}{2}\right]$$

Dividend Discount Models

We now seek expressions for the equity price using discounted dividends, at some rate r. Unfortunately, we cannot sum the dividends analytically. However, if h is reasonably large, we can neglect terms in A^h and hence obtain the approximation:

$$\mathbf{E}_t(D_{t+h}) \approx D_t \exp\left[\left(\mu + \frac{\sigma_x^2 + \sigma_y^2}{2}\right)h + \sigma_x \frac{Ax_t - e_t - (1+A)\frac{\sigma_x}{2}}{1-A}\right]$$

Now summing, and denoting the yield by Y_t, the price is given by

$$\begin{aligned}
\frac{D_t}{Y_t} &= \sum_{h=1}^{\infty} \exp(-rh)\mathbf{E}_t(D_{t+h}) \\
&\approx D_t \sum_{h=1}^{\infty} \exp\left[-\left(r - \mu - \frac{\sigma_x^2 + \sigma_y^2}{2}\right)h + \sigma_x \frac{Ax_t - e_t - (1+A)\frac{\sigma_x}{2}}{1-A}\right] \\
&= D_t \frac{\exp\left[-\left(r - \mu - \frac{\sigma_x^2 + \sigma_y^2}{2}\right) + \sigma_x \frac{Ax_t - e_t - (1+A)\frac{\sigma_x}{2}}{1-A}\right]}{1 - \exp\left[-\left(r - \mu - \frac{\sigma_x^2 + \sigma_y^2}{2}\right)\right]}
\end{aligned}$$

then the yield is:

$$Y_t \approx \left[\exp\left(r - \mu - \frac{\sigma_x^2 + \sigma_y^2}{2}\right) - 1\right] \exp\left[-\sigma_x \frac{Ax_t - e_t - (1+A)\frac{\sigma_x}{2}}{1-A}\right]$$

Price Changes

We can consider the distribution of price changes; we have

$$\begin{aligned}
&\log\left(\frac{D_t}{Y_t} \div \frac{D_{t-1}}{Y_{t-1}}\right) \\
&\approx \exp\left[\mu + \sigma_x\left(x_t + \frac{Ax_t - e_t - (1+A)\frac{\sigma_x}{2}}{1-A} - \frac{Ax_{t-1} - e_{t-1} - (1+A)\frac{\sigma_x}{2}}{1-A}\right) + \sigma_y y_t\right] \\
&= \exp\left[\mu - \sigma_x e_t + \sigma_y y_t\right]
\end{aligned}$$

We can see then that price changes are pure white noise – there are no white thunder terms lying around. The price changes will be correlated with dividend changes, the main correlation arising because of the σ_y term.

Behaviour of the Yield Process

Examining the yield itself, we can see that

$$\log Y_t \approx const - \sigma_x \frac{Ax_t - e_t}{1 - A}$$

We now recall that $Ax_t - e_t$ was our embedded first order autoregressive process. We can then deduce that log Y_t is also a first order autoregressive process, with covariance function:

$$\mathbf{Cov}(\log Y_t, \log Y_s) = \frac{1 + A}{1 - A} \sigma_x^2 A^{|t-s|}$$

This would allow us to calibrate both A and σ_x. We notice that the yield variability is associated entirely with the white thunder coefficient – it is unaffected by the white noise terms.

Applications to More Complex Models

Although our white noise approach to equity markets looks promising, our model so far contains a number of empirical deficiencies, including the fact that dividend volatility and price volatility are constrained to be equal. In practice, the price volatility is usually two or three times the annual dividend volatility.

One way to square this circle is to turn to more complicated dividend models. For example, if we allow the mean underlying the dividend growth to be a random process itself, we can obtain price volatility than substantially exceeds the annual volatility in dividend growth. Plainly a great deal of empirical research is needed to validate these ideas; however, initial investigations are encouraging. This may even provide a resolution of the long-standing equity volatility puzzle in finance.

More generally, the concept of white thunder can be used to enhance any time series model which depends on white noise for inputs. For example, we could take an autoregressive inflation model, and use white thunder instead of white noise to improve the accuracy of short term model predictions, while having no effect on the long term statistical properties of the model.