

Bayesian Financial Dynamic Linear Modeling in APL

Bayesian statistics is a brand-new idea that's only about 235 years old. The paper that was to immortalize the last name of the Reverend Thomas Bayes, a Nonconformist minister born in 1702, was published in 1763. Unfortunately for any fame he may have hoped to gain from what proved to be his most influential work, Bayes died some 3 years before.

From such inauspicious origins, his contribution to statistics continued inauspiciously. His ideas flourished briefly in the latter part of the 18th century, being taken up by the great mathematician Laplace, before languishing in relative obscurity until this century. Even though revived in the early part of the century, first by Ramsey, then by deFinetti, his ideas only became widespread in recent years.

So, what is this great idea that has come to us through centuries? At first glance, it may appear to be rather trivial. So trivial, in fact, we may briefly derive it here without further exposition.

Bayes examined the problem of contingent probability. We may start, simply enough, by noting that the probability of an event P, denoted here as $\text{Prob}(P)$, contingent on circumstances (or Data) D, is the product of $\text{Prob}(P)$ and $\text{Prob}(D|P)$. This latter formulation may be read "the probability of D given P". So, since

$$\text{Prob}(PD) = \text{Prob}(P) \times \text{Prob}(D|P)$$

and, the reverse being true and equivalent, that is

$$\text{Prob}(DP) = \text{Prob}(D) \times \text{Prob}(P|D)$$

Taking these two as equal to each other leads us to the common formulation of Bayes's Theorem:

$$\text{Prob}(P|D) = \frac{\text{Prob}(D|P) \times \text{Prob}(P)}{\text{Prob}(D)}$$

There might not seem to be much here, so why should this simple theorem have provoked so much contention and fallen into such conspicuous disfavor? The key here is the interpretation of the prior Data: it includes any information we see fit to include: this makes it "subjective," hence often thought to be not fit for proper scientific and mathematical treatment. That is, the powerful effect of Bayes's Theorem is to allow us to incorporate non-traditional, perhaps inexact, data into our probability formulations.

This denigration of subjectivity, in a historical context of pride in exact, "hard" science, continues to plague Bayesian techniques even up until recent times. In fact, one reference (in Kuhn's "Readings in Game Theory") to Bayesian techniques appears to justify calling it Bayesian only because of the general use of subjective probability estimations. One much-cited early work about Bayesian techniques is called "Studies in Subjective Probability" (Kyburg, 1964): at the time it was published, this was the defining feature of Bayesian statistics.

However, as we shall see, this perceived subjectivity of the technique is misconstrued as a weakness, when it is rather a strength because it acknowledges and deals with the "outside" information underlying all statistical inference. This combining of "hard" and "soft" data is helpful in the context of the Dynamic Linear Models we will explore later on.

Another strength of Bayesian inference has to do with its approach to probability, an approach often called "the inverse probability problem" because it is backwards in relation to traditional probability. This problem is stated thusly:

Given the number of times an event with unknown likelihood has occurred or failed to occur, what is the chance that the probability of it happening in a single trial lies somewhere between 2 degrees of probability?

For example, one of the Bayesians' favorite example of this is the urn problem: given an urn containing 2 colors of balls, say black and white, and a series of tests whereby we draw a ball, note its color and return it to the urn, what is the likelihood of various combinations of the 2 colors? For instance, given an urn with 5 balls and the evidence that 3 draws with replacement yield only white balls, what is the probability that all 5 are white?

This is the inverse of a typical problem in probability that would be more along the lines of: given an urn with 3 white and 2 black balls, what are the odds of drawing 3 white balls, 1 at a time with replacement? Before we delve further into the Bayesian example, notice how more commonly applicable to typical finance problems is this former sort of formulation than is the latter. A classical probability problem would be something like: given the mean returns and volatility of some assets, how are they likely to perform? However, the inverse problem would be something like: given the performance of these assets, what are their likely returns and volatilities? We can see from this that the inverse problem better matches what we often have to work with in terms of real data.

Before we look more at the urn problem, we need to extend Bayes's Theorem to multiple events. Given a series of k events P_i with associated likelihoods $\text{Prob}(P_i)$ and prior observed data D , the contingent probability of a particular event P_i given observations D is stated thusly:

$$\text{Prob}(P_i|D) = \frac{\text{Prob}(P_i) \times \text{Prob}(D|P_i)}{(\text{Prob}(D|P_1) \times \text{Prob}(P_1)) + \dots + (\text{Prob}(D|P_k) \times \text{Prob}(P_k))}$$

In APL, this is shown by the function *BayesPP*.

```

∇ R←BayesPP PROBS
[1]  a Calc Bayesian posterior probability given PROBS: 2 row mat:
[2]  a [0;]P(P1), P(P2)...; [1;]P(P1|D), P(P2|D)...; i.e. [0;] isolated
[3]  a event probability, [1;] conditional probability given data D.
[4]  R←R÷+/R←×+PROBS
∇

```

In practice, this works as follows. Suppose we have an urn with 5 black and white balls in unknown proportion. We draw 3, one at a time with replacement; all are white. What is the probability that all 5 are white? Invoking the APL function

UrnProbAllWhite 5 3

gives us the numerical answer of 0.15625. How do we work this out using the above theorem? We must solve for the probability of all the balls being white given the evidence of 3 white draws ($\text{Prob}(P_i|D)$). The text of this function and a discussion follows.

```

▽ P←UrnProbAllWhite NNW;□IO
[1]   a Bayesian urn problem: given NNW[0] (ostensibly black and white) balls
[2]   a and experimental evidence that NNW[1] white ones picked, with
[3]   a replacement each time, what is probability that all are white?
[4]   □IO←0
[5]   P←((1NNW[0]+1)÷NNW[0])*NNW[1]           a 0 to N possible balls of color
[6]   a picked → chance of picking NW white ones for each proportion possible.
[7]   P←+/P×(BinomialCoeff NNW[0])÷2*NNW[0] a 2* because 2 possible colors;
[8]   a would have to use other than binomial expansion if more than 2.
[9]   a P is now all probabilities of all combinations. Applying Bayes's
[10]  a theorem: (Prob(all selections white given all balls in urn are white ×
[11]  a Prob(all balls in urn are white)) ÷ Sum of Prob(all combos)
[12]  P←(((1÷2)NNW[0])*3)×2*NNW[0])÷P
▽

```

The subfunction *BinomialCoeff* is defined thusly:

```

▽ R←BinomialCoeff N;□IO
[1]   a Give coefficients of Nth binomial expansion.
[2]   □IO←0  ◊ R←((1N)!N),1
▽

```

Working 1st on the divisor of the preceding equation, calculated in line 5 of the APL function, we calculate the probability of the datum D (3 white samples) for each of the possible P_i s. These latter range from no white balls (all black) to 5 white balls, or the fractions 0/5, 1/5, 2/5, 3/5, 4/5, and 5/5. These are all the $\text{Prob}(P_i)$ s.

Line 7 of the APL function figures the likelihood of 3 white draws given each of the possible combinations of white and black balls. The denominator of each of these is 2*5 or 32 because this is the number of possible combinations of 2 things taken 5 at a time. The numerator is the possible arrangements of each combination, e.g. 1, 5, 10, 10, 5, and 1 corresponding to: 1 way to have all 5 black, 5 ways to have 1 white and 4 black, 10 ways to have 2 white and 3 black, 10 ways to have 3 white and 2 black, 5 ways to have 4 white and 1 black, and 1 way to have all white. Combining all the $\text{Prob}(P_i)$ s with the $\text{Prob}(D|P_i)$ s, multiplying them together and summing the results gives us (in math):

$$\left[\left(\frac{0}{5} \right)^3 \times \left(\frac{1}{32} \right) \right] + \left[\left(\frac{1}{5} \right)^3 \times \left(\frac{5}{32} \right) \right] + \left[\left(\frac{2}{5} \right)^3 \times \left(\frac{10}{32} \right) \right] + \left[\left(\frac{3}{5} \right)^3 \times \left(\frac{10}{32} \right) \right] + \left[\left(\frac{4}{5} \right)^3 \times \left(\frac{5}{32} \right) \right] + \left[\left(\frac{5}{5} \right)^3 \times \left(\frac{1}{32} \right) \right]$$

or (in APL, origin 0):

$$((((16)÷5)*3)+.×1 5 10 10 5 1÷32$$

which equals 0.2. The difference between these 2 expressions prompts me to question the reader: which of these 2 expressions looks simpler? Which do you think took about 5 minutes to enter and which took about 5 seconds?

The last line of the APL function calculates the numerator in Bayes's Theorem above, which is the probability of 3 white draws given all 5 balls being white, and divides it by the denominator to give our answer. A more general version of this is the APL function *UrnProblem* that returns all probabilities instead of just the "all-white" one. An even more interesting function would allow a third input of the number of black balls in a sample instead of just restricting the possible data to observations of all-white draws. This is left as an exercise for the reader. The text of *UrnProblem* follows.

```

∇ P←UrnProblem NNW;□IO;NB;NWhite;Pk;Num;Denom
[1]   a Bayesian urn problem: given NB (ostensibly black and white) balls
[2]   a and experimental evidence that NWhite white ones picked, with
[3]   a replacement each time, what are probabilities that 0 to NB are white?
[4]   □IO←0 ◊ NB NWhite←NNW
[5]   Pk←((1NB+1)÷NB)*NWhite                      a 0 to N possible balls of color
[6]   a picked → chance of picking NW white ones for each proportion possible.
[7]   Denom++/Num←Pk×(BinomialCoeff NB)÷2*NB a 2* because 2 possible colors;
[8]   a binomial expansion assumes prior normal distribution for 2 colors.
[9]   a Apply Bayes's theorem:
[10]  P←Num÷Denom
    ∇

```

One point in the above exposition that deserves further mention is our gloss on the use of binomial coefficients (line 7 in *UrnProblem*). This introduces a primary facet of Bayesian inference: the use of a prior distribution. By using the binomial coefficients to weight the black and white combinations, we are in fact assuming a prior distribution that approximates the normal distribution (for discrete values). The effect of this prior distribution may be seen in the values generated by

```

UrnProblem 6 1
0 0.03125 0.15625 0.3125 0.3125 0.15625 0.03125

```

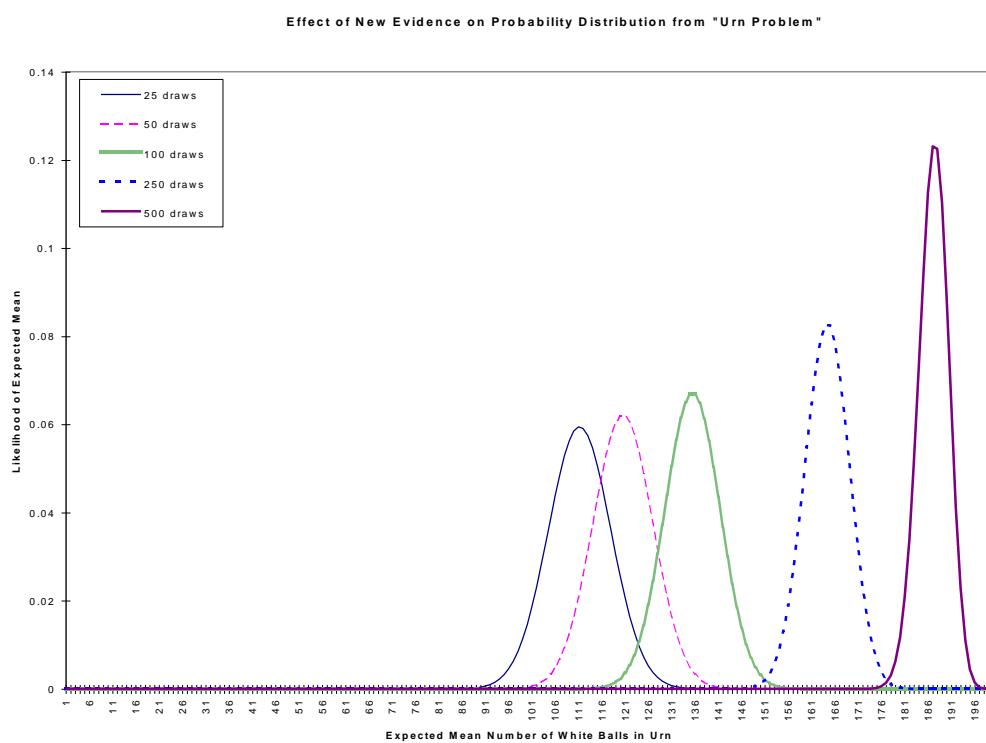
versus the result of

```

UrnProblem 6 6
0 0.0000275999 0.00441599 0.0670678 0.282623 0.431249 0.214617

```

Notice how, in the first case (for 1 draw from an urn containing 6 balls), the peak values are the 2 in the center whereas, in the second case, the peak has moved to the second-to-last value. To better see this, look at the "Evidence Effect" graph: this shows the results of *UrnProblem* run for an urn containing 200 balls and 25, 50, 100, 250 and 500 draws of white balls. Notice how the distribution shifts in the manner hinted at by our small example above. What does this mean?



In the first case, our observation of 25 white draws for 200 balls, combined with our prior (assumed) normal distribution leads us to put the most likely combinations near the center. This shows the strong influence of the prior distribution with little additional evidence. The subsequent cases, each with more white draws, move the peak toward the all-white end of likely distributions. Of course, evidence like this might lead us to question our choice of prior distribution. In any case, we can see how this Bayesian technique allows us to combine our estimate of a distribution with subsequent evidence about that distribution.

While we're on the subject of prior distributions, it's worth noting that the traditional standpoint of statistics, from which the "subjectivity" of Bayesian statistics has been attacked, is in fact explainable in Bayesian terms as one with a "flat" prior. That is, if we decide to study the frequency of occurrence of heads and tails when flipping a "fair" coin, we can either assume, as traditionalists do, that we have no information about the likely outcomes (that is, all probabilities are equal, hence the prior distribution is a straight line), or we might acknowledge that this assumption is just one of many possible prior distributions. Under the traditional assumption, after 1 trial of flipping the coin, we would have to say that the likelihood of heads is 100% (or 0%, depending on which turned up).

This influence of the prior distribution suggests a further modification to the *UrnP*roblem code: we could supply a prior distribution function as one of its arguments. This change, along with the one suggested earlier that allows for more generality in observations, would give us a version of the function that starts to appear to have more practical application than the small samples of Bayesian inference we've seen so far. However, we will leave this pursuit to interested readers. Instead, we will now look at the other important technique that, combined with Bayesian inference, can give us a robust financial forecasting system: Dynamic Linear Models.

Dynamic Linear Models

The following pair of equations characterizes a dynamic linear model:

Type	Equation	Distribution Mean, Variance
Observation	$Y_t = F_t' \theta_t + v_t$	$v_t \sim N[0, V_t]$
System	$\theta_t = G\theta_{t-1} + w_t$	$w_t \sim N[0, W_t]$

The notation in the column labeled "Distribution Mean, Variance" gives the distribution of the error term for each equation; the expression $N[m, v]$ specifies a normal distribution with mean "m" and variance "v".

The variables in these equations have the following meanings:

Variable	Description
Y_t	vector of dependent variables (e.g. asset returns)
F_t'	(transposed) matrix of explanatory factors
θ_t	regression factors
v_t	observation errors
G	state evolution matrix
θ_{t-1}	previous time's θ
w_t	system errors

The ultimate aim of a Dynamic Linear Model (DLM) is to estimate the Y values in terms of a likely mean and variance. We may think of the equation that purports to explain these Y values as a linear equation with variable coefficients applied to time-varying explanatory factors. For instance, a model of bond prices might state the observation equation using factors known to influence bond prices, such as government interest rates and the spreads of corporate rates.

The system equation may be thought of as a system of relations between the observed factors. However, the dynamism of the model derives from assuming that these relations vary.

Bayesian Forecasting with Dynamic Models

The preceding system of equations is updated by successively examining predictive factors with their corresponding returns, predicting the mean and variance of expected returns based on the next set of factors, and modifying these predictions according to the actual returns. We can see how this is analogous to our earlier, simpler problem of estimating the unknown composition of balls in an urn based on the evidence of samples from the urn. In the case of the urn, we saw how the probability distribution moved from the initial assumption of normally distributed with an equal mix of black and white to a higher likelihood of all white balls based on the evidence of repeated samples of white balls. Similarly, our forecasting model starts with assumptions about means and variances of returns then modifies these assumptions based on successive samples of correlations between econometric factors and asset returns.

The West and Harrison book (Bayesian Forecasting and Dynamic Linear Models, Springer-Verlag, 1997) provides a series of increasingly complex sets of data to model in chapter 3. Three of these examples are implemented in APL in the functions *Theorem3p1*, *Theorem3p1_1*, and *Theorem3p1_2*. The associated data sets are contained in the variables *TBL3P1*, *TBL3P2*, and *TBL3P3*, respectively. These may be of interest for use with the book. However, we will look more at a multivariate version of the model that more closely parallels, though it is somewhat simplified, the one in use at my company for asset allocation.

The model on which we will concentrate is the one embodied in the function *MultiVDLM3*. Before we look at this, though, a word about the data used in this example. To avoid problems with proprietary or purchased data, we'll be using generated datasets. These are outputs from the *GENFACRETS* function. Hence, a brief look at this function is in order.

```

∇ FR←GENFACRETS NUM;◻IO;MAX;FACS;RETS
[1]   a Generate a set of factors and returns associated with them.
[2]   ◻IO←0 ◇ FR←(0 4p0)(0p0) ◇ MAX←1E9
[3]   a FACS←-\1 2 1 2o1 2 3 4÷5 6 7 8
[4]   FACS←(5+?4p100)|1+?4p1E9
[5]   RETS←-/FACS
[6]
[7]   L_NEXT:FACS←1 2 1 2o0.5 1 1.5 2×FACS+2 3 4 5
[8]   FACS←FACS+0.01×+/-5+?(pFACS),20)p11 a Handful of little normal noise
[9]   RETS←-/FACS
[10]  RETS←RETS+0.01×+/-5+?(pRETS),20)p11 a Handful of little normal noise
[11]  →(NUM>1p0>FR←FR,[0]“FACS RETS)pL_NEXT
∇

```

Notice how the factors are generated using a “hidden” function and the returns are based on these factors. A typical use of this function might look like this:

```
Factors Responses←GENFACRETS 120
```

to assign the 2 vectors *Factors* and *Responses*, each with 120 elements as specified by the function's right argument.

The accompanying illustrations of forecasts compared to "actual" data use these generated data as inputs to the multivariate dynamic linear model. Looking now at the primary instance of this model, the APL function *MultiVDLM3*, there are a few important points to note. A full explanation of this model is beyond the scope of this paper; however, one of the books by West and Harrison or Pole, West, and Harrison, should provide sufficient detail (see the bibliography). Bearing this in mind, let us examine the few points on which we will concentrate to illuminate the robustness of this type of model. The text of the function follows (next page).

```

▽ DMO<-MultiVDLM3 FGVWYMC;IO;Wt;Vt;MC;SZ;CT;mp;Cp;F;Y;G;Ft;Yt;Rt;ft;Qt;
At;at;et;mt;Ct;DELTA;nt;np;St;Sp;TIT
[1]   a Multivariate Dynamic Linear Model, adapted from West and Harrison.
[2]   a FGVWYMC: [0] Factors; [1] G (evolution matrix); [2] V (observation
[3]   a variance); [3] W (system variance); [4] responses (Y values to
[4]   a predict); [5] MC: prior mean and variance; [6] (optional) delta
[5]   a (discount factor). E.G. for 3 factors predicting 1 return:
[6]   a MultiVDLM3 (10 3p130) (3 3p4t1) (30 1p2x130) (1 1p1) (3 3p0.01)
[7]   a (15 9) (0.6)
[8]   a OR DMO<-MultiVDLM3 F ((2p-1pF)p(1+pF)t1) (,COVM R) (COVM F)
[9]   a ((pR,1)pR) ((AVG R) (STDDEV R)) (1)
[10]  a DMO[1;]: [;0] Forecast mean and [;1] variance, [;2] posterior system
[11]  a (theta) mean and [;4] variance, [;5] adaptive coefficient, [;6] error,
[12]  a [;7] Scaling factor.
[13]
[14]  CT<-IO<-0 ◊ TIT<-ft 'Qt' 'mt' 'Ct' 'At' 'et' 'St'
[15]  F G Vt Wt Y MC DELTA<-7↑FGVWYMC,1 ◊ SZ<-1p0F ◊ DELTA<-DELTA*0.5
[16]  DMO<-(pTIT)p<0.1p0 ◊ np<-Sp<1
[17]  mp Cp<-(pWt)p"MC a Prior mean and variance: coerce shapes if scalar
[18]  at<-mp
[19]
[20]  L_D01: Ft<-qF[,CT;] ◊ Yt<-Y[,CT;]      a 1 row mats for conformability
[21]  at<-G+.xmp
[22]  Wt<-(DELTAxG+.xCp+.xqGxDELTA)-G+.xCp+.xqG a Discounting
[23]  Rt<-(G+.xCp+.xqG)+Wt      a Prior at time t: theta variance
[24]  ft<-(qFt)+.xat      a 1-step forecast -means
[25]  Qt<-Sp+(qFt)+.xRt+.xFt      a and -variance for forecast Y
[26]  et<-Yt-ft      a 1-step error forecast
[27]  nt<-np+1      a Start on posteriors
[28]  At<-Rt+.xFt+.xZQt      a Adaptive coefficients,
[29]  mt<-at+At+.xet      a Posterior thetas' means
[30]  Ct<-Rt-(qAt)+.xAt+.xQt      a and (scale-free) variances
[31]  St<-'pSp+((qet)+.x(ZQt)+.xet)-1)xSp:nt a Scaling factor (p. 110 (d))
[32]  Ct<-(St:Sp)xCt
[33]  DMO<-DMO,[0]~>ft Qt mt Ct(qAt)et St
[34]  Cp mp Sp np<-Ct mt St nt      a Current to previous for next loop
[35]  →(SZ>CT<-CT+1)pL_D01 ◊ DMO<-TIT,[-0.5]DMO
▽

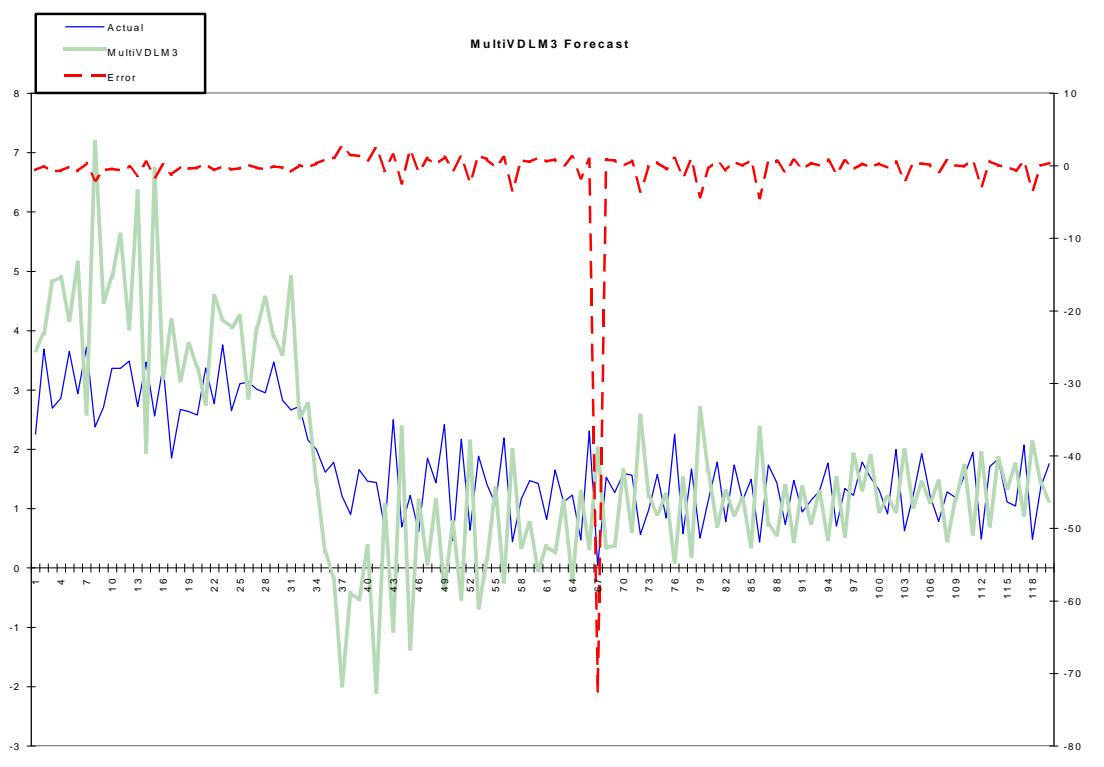
```

First, let's dispose of variables set to rather trivial values for this example: G and $DELTA$. We set G to be the identity matrix: this corresponds to a random system evolution. We set $DELTA$ to be 1; this might have a value slightly less than one to more greatly discount the effects of data, which is further in the past. In addition to these 2 settings, made uninteresting for simplicity, we also take a simple default on the prior mean and variance setting ($FGVWYMC[5]$ in the function) by using the actual mean and variance of the return series. Disposing of all these leaves us to consider only the factors and returns, which we discussed above, and the 2 variances W and V , for the system and observation series.

How to deal with these 2 inputs remains a subject of some research. We'll look mainly at V , the observation variance, to illustrate the robustness of DLMs and how well the Bayesian framework allows us to incorporate new prior information.

Many of the models (in West and Harrison, 1997) assume that this observation variance is known. In fact, it is tempting to use the same shortcut we used for the prior mean and variance values by essentially "looking into the future" and basing the value on our present data. However, this detracts from the ability of the model to adapt to new information by reinforcing the values of past data. However, notice that this is essentially what we did in our example just to get some crude numbers to look at. The important thing to notice is that the system of equations given above specifies V as varying according to time t but our APL implementation treats it as a known constant.

This limitation of our sample implementation detracts from its forecasting power. A better model would attempt to forecast the variance series as it does the return series and progressively correct these forecasts with the receipt of new data. Indeed, the flexibility of this sort of model allows us to do even better than this by incorporating anticipated future data. Since we do not have such an implementation in our simple example, let's see how and why we might do this. First, however, let's look at an example of the function's forecast on some data "made up" by the *GENFACTETS* function as shown above. These results are shown graphically in the following chart.



Notice how the earlier attempts at forecasting, on the left, diverge widely because the model hasn't "learned" much about the relation between the factors and their associated responses. However, it does track the change in level starting at about points 31 to 43, even though it initially overshoots. Suppose that this change in level were in response to events we could anticipate to some extent: we know that the fundamentals affecting our responses are about to change but we can't quantify them exactly.

Say, for example, that one of the assets whose return and variance we are forecasting is the Hong Kong dollar. Since the value of the Hong Kong dollar has been fixed to that of the U.S. dollar for many years, our model will build a misleading picture of this currency's behavior based on historical data. This is a very pertinent issue at the time I write this, since that colony has recently reverted to Chinese rule, Asian markets in general have become more volatile of late, and there are indications that the end of this fixed rate regime is close at hand. According to an article in Barron's (the column "The Striking Price", November 17, 1997), the American stock exchange "plans to revise its Hong Kong Option Index so that its value is calculated using a floating rate of exchange for the Hong Kong dollar rather than the fixed value that's now used."

These factors make it very likely that a Hong Kong dollar model will change behavior drastically in the near future. In addition, this sort of deliberate change will usually be announced in advance, thus aiding the modeler using Bayesian DLMs. A more complete model than the one presented here would use a true time-varying variance and would implement it in such a way as to incorporate an external multiplier. Though one might hazard a guess as to the direction of the Hong Kong dollar when it begins to float, one is not obligated to.

Though a fuller system than the one outlined here might also accommodate external inputs to the calculations of means as well as variances, this would require a view on the market that we might not be prepared to risk. However, it does not require much of a leap of faith to predict that the volatility of the Hong Kong currency will increase relative to the U.S. dollar some time soon.

A Bayesian DLM naturally and easily incorporates such new information, even when it is rather vague and subjective as such advance information often is. We are required to assign some value to our variance multiplier, but, in doing so, we can use whatever other information we see fit (and in this example there is a strong argument for a certain lower limit.) This ability to seamlessly incorporate judgment based on new developments makes these kind of models very attractive, especially since it minimizes or eliminates re-programming and allows vague information to be introduced in a well-controlled manner.

This document is for informational purposes only. Opinions expressed are our present opinions only, and are subject to change. In preparing this presentation, we have obtained information from sources we believe to be reliable, but do not offer any guarantees as to its accuracy or completeness.