MOODY'S ANALYTICS

RESEARCH
JUNE 2015
INSURANCE ERS

Economic Forecast Validation: Evaluating the Calibration of Models for Interest Rates

Aubrey Clayton PhD
Steven Morrison PhD
Moody's Analytics Research

Contact Us

Americas
+1.212.553.1658
clientservices@moodys.com

Europe
+44.20.7772.5454
clientservices.emea@moodys.com

Asia (Excluding Japan)
+85 2 2916 1121
clientservices.asia@moodys.com

Japan
+81 3 5408 4100
clientservices.japan@moodys.com

Overview

In this paper we consider a framework for evaluating real-world probabilistic forecasts of economic variables, particularly nominal interest rates over quarterly time horizons. The main questions we seek to answer are:

(1) How reliably does the model predict the true frequency of observed interest rates?
(2) How well does the model distinguish between forecast distributions under changing conditions, e.g., high and low interest rate volatility?

After a brief survey of the analytical tools, we evaluate the Moody's Analytics Economic Scenario Generator standard models for projecting nominal interest rates, as most commonly used in short term asset-liability management or calculating 1-year Value-at-Risk.

Results indicate that the model predictions are reliably in agreement with observed rates and outperform both a "Through-the-Cycle" calibration and a "Stationary" calibration in capturing near-term fluctuations due to changes in volatility. Some refinements in individual and joint distributional relationships are proposed.

Contents

1. Introduction

The setting we consider in this note is the projection of financial and economic variables, particularly nominal Government bond yields, for real-world purposes. We are primarily concerned with forecasts described in probabilities, for example statements of the type, "The probability of the 10-year Treasury rate one quarter from now exceeding x is y." Or, considering all possible exceedance levels together and describing the probabilities by means of a parametric distribution, we will consider statements such as, "The probability distribution of the 10-year rate in one quarter is log-normal with parameters μ and σ," etc. The source of the forecasts will be the Moody's Analytics Economic Scenario Generator (ESG), calibrated quarterly to project economic variables over periods of one year or less.

The problem of rigorously evaluating such models is a difficult one with much inherent room for subjectivity. In contrast to the exercise of projecting variables on a market consistent or risk-neutral basis, in which the model's ability to replicate observable market prices is a final check of how correctly it has been calibrated, projecting a distribution for real-world behavior opens up the discussion to many potentially uncomfortable questions:

- How do we know our probabilities (either the distributional types we use or the particular parameters we populate them with) are the correct ones?
- Were particular historical events captured adequately in our projected distributions for that time period? For example, at what percentile level would we have placed the extreme financial turmoil seen at the end of 2008, and was this correct?
- On what information are our forecasts conditioned, and how well do our forecasts change in different market environments or as a result of new information?
- How much confidence do we put around the stability of our forecast distributions from forecast to forecast? That is, are we attempting to capture "risk" in the sense of *known* expected variability or "uncertainty" in the Knightian[1] sense of possibly *unknown* variability?
- How do we account for joint behavior in projecting multiple variables simultaneously? For example, how can we tell if we've assumed the correct correlation between two variables?

There is perhaps no one right answer to any of these questions, and we should likely expect the discussions to continue indefinitely. Nevertheless, as long as we continue to produce calibrations of real-world models or support others who do, we should attempt to gain greater degrees of confidence in the correctness and usability of those models and to improve upon them wherever possible.

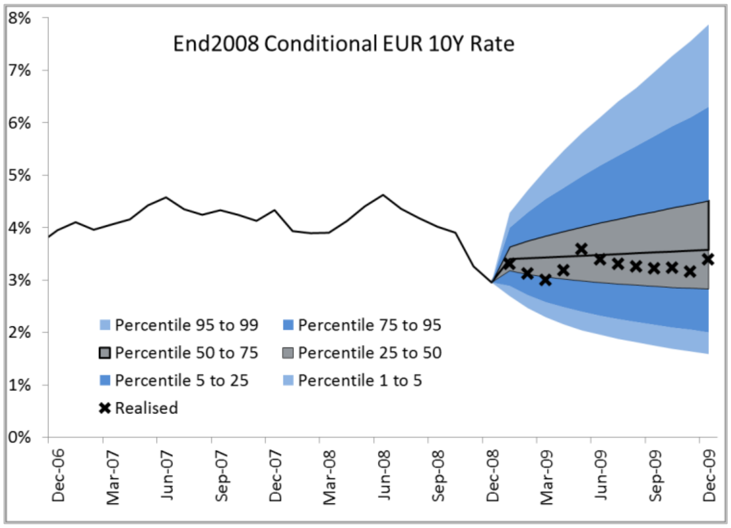
Previous attempts to historically validate, or "back-test," standard calibrations of the ESG models have been limited in scope primarily to equity market returns where more data is available and the situation is perhaps better understood.[2] The problem is made especially difficult in the context of projecting interest rates since we are constrained by the availability of data and of forecasts, and we are often forced to consider pooling data from multiple economies and from multiple disparate time periods. Some effort has been made towards evaluating interest rate distributions, but on a relatively narrow basis.[3] In each instance, the main analysis has been a comparison of realized observations against a projected distribution for a single variable and a judgment about whether the observations appear reasonable. The figure below shows a typical example.

[1] Knight (1921)

[2] See El Cherif (2010).

[3] For example, see Conn *et al.* (2013)

Figure 1: Example Back-Testing Validation Chart



Whether the realized values appear "reasonable" is open to interpretation and difficult to quantify.

Our goal is to extend the previous analysis in two ways: by considering a larger number of forecasts for joint distributions across multiple economies, and by adding some quantitative assessment of the correctness of the forecast distributions. We compare each forecast with appropriately normalized historical data and score the results both on the agreement between projected distributions vs. realized frequencies and on the model's ability to distinguish different economic conditions, i.e., periods of high and low interest rate volatility.

The rest of the note is organized as follows:

Section 2 gives an outline of our forecast validation methodology.

Section 3 reviews the calibrations for the Moody's Analytics interest rate models.

Section 4 shows the results of applying the forecast validation methods to the interest rate forecasts and compares with other possible model calibration approaches.

Section 5 concludes with suggestions for model improvements and further research.

2. Methodology

Throughout this discussion, we use the word "forecast" generally to mean any statement about future events with a quantifiable and measurable outcome. The simplest kind of forecast to evaluate is a deterministic binary statement about the occurrence of a particular event, e.g., "The 10-year Treasury rate one quarter from now will be higher than its current value," since there will presumably be an opportunity to evaluate the correctness of the forecast by waiting until the appointed time and recording whether the event in question occurred. Results over repeated observations can then be understood by summary statistics such as the "hit rate" and "false alarm rate,"[4] and these concepts are easily generalized to multi-category predictions such as whether a certain variable will fall into one of a number of pre-defined states.

However, we will consider *probabilistic* forecasts of the type, "The probability of the 10-year Treasury rate one quarter from now exceeding x is y" made via Monte Carlo simulation of economic variables in the ESG. Though this example is phrased as the probability of a binary outcome, it easily extends to probabilistic forecasts for a continuous variable, e.g., $X =$ "the interest rate one quarter from today," by considering binary events such as "The interest rate will exceed x" for different values of x. Clearly, specifying forecasts for all of the latter probabilities is equivalent to giving the cumulative distribution function of the variable, $F(x) \coloneqq P[X \leq x]$. Furthermore, we consider the problem of projecting multiple interest rates simultaneously, e.g., interest rates in multiple economies, for which the forecasting probabilities are completely captured by the joint distribution of the components $x_1, \dots, x_n$.

[4]See Jolliffe and Stephenson (2012) for background.

In each case, we are interested in evaluating the "correctness" of the forecast system; however, the probabilistic nature of the forecasts makes it unclear how exactly to conduct this evaluation, even when dealing with simple binary events. For example, if a forecast calls for a 50% chance of interest rates rising and they do, in fact, rise, how do we assess whether the forecast was correct?

One possible approach would be to aggregate together many such forecasts and compare the realized frequency of occurrence with the forecast probability. For example, if, over a collection of 40 quarters in which the forecast probability of rates rising was 50% each quarter, the interest rate was actually seen to increase about 20 times, we may be inclined to say the forecasts were behaving correctly. There are at least a few deficiencies with this as a sole means of evaluating forecast quality, though:

1. Practically speaking, it may not be possible to group together forecasts with the same probability, because the probabilities themselves may be continuously varying (or simply expressed to such a precision that they effectively never repeat). How would we handle events with probability 49% vs. 51% and so on? This will obviously depend on the means by which the forecasts are being produced, but we should certainly allow for the existence of complex models that can produce probabilistic forecasts with a high degree of precision, and we should guard against our analysis being sensitive to the ways those forecasts are binned together, if possible.

2. It has not been made clear in the above description what information the various forecasts are *conditioned on*, and whether we can treat each forecast/realization pair as being independent. For example, a quarterly forecast for multiple economies may include a distribution for each one's interest rates, but if they are linked by a large amount of international trade or currency exchange, it would be natural to expect the results (and their deviations from the forecast) not to be independent. Aggregating the economies together effectively ignores this joint dependence and perhaps punishes the forecast unfairly.

3. Relatedly, comparing only realized frequencies against forecast probabilities does nothing to discourage the forecaster from reporting the same probability for every forecast, so long as this value is close to the long-run average that will be realized *eventually*. By analogy, a climate scientist whose only knowledge about our local area is that over the next several years the proportion of rainy days will likely be around 20% might be justified in forecasting a 20% chance of rain *every day*, and the above methodology (grouping together all days with the same forecast probability and comparing with the observed frequency), would seem to confirm this as an accurate forecast. However, this kind of forecast is practically useless as a measure of the chance of rain *each day*, and ideally, a probabilistic forecast should include more local variability specific to each day's conditions.

The situation with continuous variables is even more complex, since, as we've identified, their forecast distributions are theoretically equivalent to a continuum of binary probabilistic forecasts of the type "The probability of X exceeding x is y," but how are we to evaluate all of the infinite number of these forecasts simultaneously? In particular, our evaluation should account somehow for the *size* of the errors in x; for example, if we claim a 100% chance of the interest rates being less than 2%, we should be penalized more if the realized value is 6% instead of 2.01%. At first glance, measuring realized against predicted frequencies does not appear to differentiate between the two outcomes, since both are predicted to have 0% probability.

We will address the concerns in points 1 and 2 in the next section by means of standard statistical methods designed to test consistency between a set of realized data and a set of probabilistic forecasts, both in terms of their marginal and their joint distributions. Point 3, regarding the ability of the forecasts to *resolve* different conditions, is more subtle and requires an additional analytical framework, which we take up in the discussion of scoring rules in section 2.2.

2.1 Statistical Methods

If, as a special case, the forecast distribution F for a continuous variable such as the 10-year interest rate were constant from forecast to forecast, then the question of agreement between forecast distributions and realized frequencies would amount to testing whether the observed interest rates $x_1, \ldots, x_N$ could be said to have been sampled from this theoretical distribution. There are many statistical tools designed for just this purpose, most notably the Kolmogorov-Smirnov test on the statistic:[5]

$$D_N := \sup_x |F^{(N)}(x) - F(x)|$$

where $F^{(N)}$ is the empirical distribution function of the data:

[5] As well as Cramer-von Mises and Anderson-Darling tests, among others.

$$F^{(N)}(x) := \frac{1}{N}\sum_{i=1}^{N} I(x_i \leq x)$$

However, in our examples the forecast distributions for interest rates $F_1, \dots, F_N$ are *not* constant; they change from quarter to quarter in order to account for different conditions when probabilities of particular interest rate movements should be expected to be higher or lower, conditional on the available information. A way of extending these methods to cover the latter situation is to normalize the realized variables via the *probability integral transform*:

$$\tilde{x}_i := F_i(x_i)$$

Under the assumption that x_i is (independently) drawn from F_i for all i, it will follow that $\tilde{x}_1, \dots, \tilde{x}_N$ are independent with uniform(0,1) distribution, so the statistical tests can be applied. For example, Hu, Levy, and Zhang (2013) used this method to evaluate credit default distribution forecasts in the Moody's RiskFrontier model.[6]

Rosenblatt (1952) considered a multivariate version of the probability integral transform, in order to test the hypothesis that a sequence of k-dimensional samples of a vector-valued $\boldsymbol{x}$ are drawn from a given distribution $\boldsymbol{F}$. Here, the transform is applied on each component of the data via the distribution function conditioned on all of the previous components:

$$\begin{aligned}
&z_1 := F_1(x_1) \\
&z_2 := F_2(x_2|x_1) \\
&\dots \\
&z_k = F_k(x_k|x_1, \dots, x_{k-1})
\end{aligned}$$

where $(x_1, \dots, x_k)$ are the components of $\boldsymbol{x}$, and F_i is the marginal distribution of x_i conditional on $x_1, \dots, x_{i-1}$.

Under the assumption that $\boldsymbol{x}$ is drawn from $\boldsymbol{F}$, we will have that $\boldsymbol{z} := (z_1, \dots, z_k)$ is a vector with a uniform distribution on the hypercube $[0,1]^k$. [For a practical application, involving multivariate forecasts of foreign exchange rates, see Diebold, Hahn, and Tay (1999).]

In greatest generality, then, we may consider a sequence of N vector-valued samples $\boldsymbol{x_1}, \dots, \boldsymbol{x_N}$, of the interest rates in k different economies, each tested against a *different* multivariate distribution $\boldsymbol{F_1}, \dots, \boldsymbol{F_N}$. For each data point we apply the multivariate probability integral transform for its particular forecast distribution, which under the given assumptions will result in a uniformly distributed k-vector.

We then have a choice in how to pool data together: in particular, we may test each k-vector $\boldsymbol{z}$ against the uniform distribution on $[0,1]^k$ using multivariate versions of the above statistics (i.e., testing both the marginal distributions and the joint distributions of the components) or we may pool the transformed components $z_1, \dots, z_k$ together and test against the uniform distribution on $[0,1]$ (i.e., assuming the independence and testing the marginal distributions only). Which approach we pick will depend practically on the available data and will naturally involve a trade-off between statistical tests that are underpowered and ones that provide less useful information.

At one extreme, given that we have N samples of k-dimensional data, after applying the multivariate probability integral transform (with different distribution for each vector) we may pool all components of all vectors together and test the hypothesis that the $N * k$ resultant values are drawn from the uniform(0,1) distribution. This test will have the greatest possible power, but if we reject the null hypothesis at some chosen level, we will not know what the source of the failure is: it could potentially be that the marginal distributions of the components were systematically biased, or that the joint distributions were incorrect, etc. At the other extreme, we can leave the k-vectors intact and test the hypothesis that these N vectors are uniformly distributed on the hypercube $[0,1]^k$. However, for even relatively large k this test will likely be underpowered unless we have many samples.

An additionally troubling possibility is that by pooling data together we may end up with a transformed data set that is apparently uniformly distributed but only as a result of offsetting biases in different forecasts. For example, if the forecast for $\boldsymbol{x_1}$ (interest rates over quarter 1) had a systematic downward bias for all economies and the forecast for $\boldsymbol{x_2}$ (interest rates over quarter 2) were upwardly biased in all economies, after transforming and pooling the data we may no longer be able to detect either bias, as the effects would cancel.

However, we note that these issues are already present in the univariate case of the transformed data $\tilde{x}_1, \dots, \tilde{x}_N$. The null hypothesis includes the assumption that the transformed samples are independent, meaning that the forecast distribution for, say, x_2 has already incorporated all relevant information available from x_1, etc. But if we reject this hypothesis we do not know whether we're rejecting the distribution assumptions themselves – and, if so, which one(s) – or the independence assumption.

[6] For additional examples of validating credit measures via similar techniques to what we have described here, see Dwyer and Korablev (2007) and Crossen, Qu, and Zhang (2011).

Furthermore, by normalizing the data and pooling the samples together we have potentially obscured any biases specific to individual forecasts.

When dealing with transforming multivariate data via the sequential conditional distribution approach described above, we also have the issue that the transformation will depend on the order in which we write the economies. Conditioning in a different order may produce better or worse looking results for the statistical tests. One natural idea mentioned in Rosenblatt (1952) is to consider all $k!$ permutations and compute the best and worst test statistics. We will examine this in more detail below, in which considering different permutations of the economies can shed some light on which joint relationships might be problematic.

In light of the above concerns, passing statistical tests on the transformed and pooled data may best be thought of as a *necessary* but not *sufficient* condition for the forecasts to be reliable, in the sense of realized frequencies being consistent with forecast probabilities. The tests may indicate ways in which our forecast distributions are systematically biased, either marginally or in joint distributions, but some kinds of bias might be obscured or undetected depending on how the data are organized, transformed, or aggregated together.

Furthermore, as described above, agreement between forecast distributions and realized frequencies is only one desirable feature for forecasts to have but is the only one that these statistical methods address.

2.2 Scoring Rules

To further assess the quality of probabilistic forecasts, we will use a continuous *scoring rule*, which will be a generalized form of the "Brier Score" introduced in Brier (1950):[7]

$$\widehat{BS} = \frac{1}{N}\sum_{i=1}^{N}(p_i - I_i)^2$$

Where N is the total number of forecasts of a particular event, p_i is the probability of the event according to each forecast, and I_i is the indicator function taking the value 1 if the event occurs and 0 otherwise. Lower values of the Brier Score are better, and the best possible forecast system ($BS = 0$) is the one that produces perfect deterministic forecasts, with $p_i = I_i$ for all i.

An advantage of the Brier Score – as opposed to other scoring rules that asymmetrically punish, say, false negatives more than false positives, is that it encourages the forecaster to apply his or her actual best judgment in constructing the forecast, since under best estimate probabilities (e.g., for a single event with probability q) the expected score is:

$$E[BS] = q * (p-1)^2 + (1-q) * p^2 = p^2 - 2pq + q$$

and this is uniquely minimized as a function of p for $p = q$. Thus, when evaluated using the Brier Score, the forecaster has an incentive not to distort or bias the forecast to report any probability other than the best estimate. Scoring rules with this property are called "strictly proper."[8]

By considering the forecast probability itself as a random variable and conditioning on its value, Murphy (1973) decomposed the Brier Score as an estimate of three components:

$$BS = E_p\left[\left(p - f(p)\right)^2\right] - E_p\left[\left(s - f(p)\right)^2\right] + s(1-s)$$

where $f(p)$ is the realized frequency of occurrence of the event given that the forecast probability was p, s is the long-term average rate of occurrence, and the expected values are taken over the distribution of p.[9] The three terms can then be further interpreted as follows:

- The first term, called the "*reliability*," compares forecast probability vs. observed frequency as in the above discussion; a perfectly reliable forecast will have a value of 0.
- The second term, called "*resolution*," records the variance of the frequencies realized over the various forecasts, $Var_p[f(p)]$. Generally speaking, good forecasting systems should have realized frequencies that vary significantly around the mean value s, since this indicates that the forecasts have incorporated information that distinguishes different

[7] The version presented here differs slightly from Brier's original proposal but is the more commonly used formulation now.

[8] See Jolliffe and Stephenson (2012), chapter 2.

[9] Note that the quantity $\widehat{BS}$ is the in-sample estimate of this theoretical value, and so the relationship will only hold in the large N limit.

environments from each other, i.e., periods when the event is expected to occur more or less frequently.[10] This may be true even if the observed frequencies are not themselves equal to the forecasts; for example, a forecaster who predicts a 10% chance of rain *if and only if* it actually rains the next day has done a remarkably good job of predicting the weather, even though it may not seem that way!

- The final term, called "*uncertainty*," is independent of the forecasting system and encodes the overall variance of the observations (i.e., Bernoulli random variable with success probability s); events with a baseline occurrence rate around $s = 0.5$ will naturally be harder to construct quality forecasts for than extremely rare events.

Seen in this way, it's apparent that the previous methods of validating economic models by comparing realized against predicted frequencies have only dealt with the reliability and not the resolution term. In particular, the "climatological" forecast that always issues the same forecast probability s is perfectly reliable but has the worst possible resolution.

For a continuous variable falling into bins demarcated by values $y_1 < y_2 < \cdots < y_K$ with associated (cumulative) forecast probabilities $p_1 := P[x \leq y_1], \ldots, p_K \coloneqq P[x \leq y_K]$, the Brier Score can be generalized to the "Ranked Probability Score"[11]

$$RPS = E\left[\frac{1}{K}\left(p_k - I(x \leq y_k)\right)^2\right]$$

To extend this to forecast distributions of a continuous variable, Matheson and Winkler (1976) defined the "Continuous Ranked Probability Score"

$$CRPS = E\left[\int (F(x) - I(x_0 \leq x))^2 dx\right]$$

Here, the expected value is taken over the distribution of *possible* forecast distributions F, each with a corresponding realized value x_0. In practice, this is estimated by the sample average:

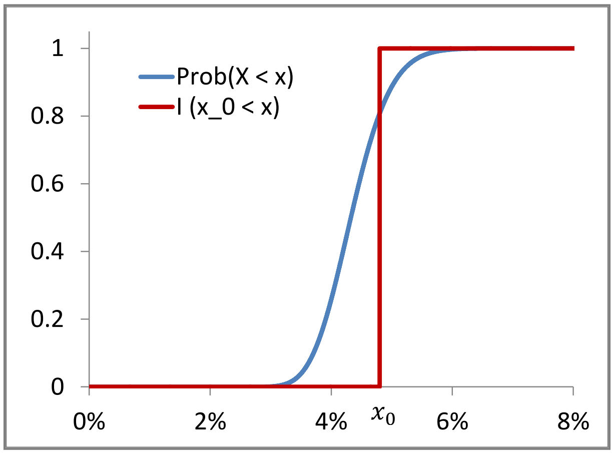
$$\widehat{CRPS} = \frac{1}{N}\sum_{i=1}^{N}\int (F_i(x) - I(x_i \leq x))^2 dx$$

Essentially, the CRPS measures the (L^2-) distance between the forecast c.d.f. and the empirical c.d.f. (Heaviside step function) for each data point and then averages over the sample, with a perfect deterministic forecast attaining the best possible value of 0. The figure below illustrates an example of the two distribution functions being compared in the CRPS calculation:

[10] To quantify the forecast resolution in a slightly different way, referred to as "sharpness," Murphy and Winkler (1992) proposed using the average of the Shannon-Weaver entropy of the forecast distributions through time, with a smaller value indicating greater sharpness, since forecasts that are closer to being deterministic have a lower entropy. This differs conceptually from the Brier Score term in that the entropy depends only on the forecast distributions themselves and not on the realized frequencies. However, if these are equal (i.e., if the forecast system is reliable), the two definitions are consistent.

[11] See Epstein (1969).

Figure 2: Distribution Functions for CRPS Calculation



For a single forecast of a binary event, represented by the value $x_0 = 1$ if the event occurs and $x_0 = 0$ otherwise, we have $F(x) = 0$ for $x < 0$, $F(x) = (1-q)$ for $0 \leq x < 1$, and $F(x) = 1$ for $1 \leq x$, where q is the forecast probability of the event. So the CRPS reduces to the Brier Score, $\left(q - I_{\{x_0=1\}}\right)^2$.

Like the Brier Score, the CRPS is strictly proper,[12] and we can see that it encodes the same kind of information as the Brier Score; namely, inaccurate forecasts will be penalized for not capturing the realized distributions of x, and forecasts with inadequate resolution will be penalized for deviating from a more deterministic forecast that consistently puts the realized value x_i in a region of high probability.[13]

These score assessments become most useful in comparing the scores of different forecasting systems. To that end, the score is commonly expressed relative to some baseline reference forecasting system, often a constant "climatological" forecast as described above. The relative improvement is known as the "Continuous Ranked Probability Skill Score" (higher values better, with a maximum of 100%):

$$CRPSS = 1 - \frac{CRPS}{CRPS_{reference}}$$

Example: Flipping Colored Coins

For a simple example to illustrate the use of skill scores, imagine a large bag full of coins, half colored red and the other half green, and suppose it is known that the red coins through some physical asymmetry have a 90% probability of coming up heads and that the green coins have a 10% probability of coming up heads. Suppose now that we begin drawing coins out of the bag at random and flipping them, and that we have two observers, A and B, constructing probabilistic forecasts of the outcome of the next toss. They are both allowed to see the coin before we flip it, but observer A is *red-green colorblind* and cannot tell which color of coin we have pulled out, whereas observer B *can* see the color.

Observer A is therefore led to forecast a constant probability of $(0.5) * (0.9 + 0.1) = 0.5$ for each coin flip, since he knows only that the chances are equal that the coin is red or green. Observer B, however, will forecast a probability of 0.9 for the i^{th} toss if the coin is red and 0.1 if the coin is green. Note that both forecast systems are perfectly reliable (they predict the correct frequencies of occurrence that will be obtained in the long-run), but we expect B's forecasts to have better resolution, since they are conditioned on more specific information about each toss. Effectively, A is playing the role of the "climatologist" who can *only* predict the long-run average frequency.

Observer A's skill score (realized over a sufficiently large number of tosses), then, will be

$$E[(0.5 - I_{Heads})^2] = (0.5) * (0.5 - 1)^2 + (0.5) * (0.5 - 0)^2 = 0.25$$

whereas observer B's skill score will be

[12] See Matheson and Winkler (1976).

[13] For a more explicit decomposition of the CRPS into reliability and resolution components, see Hersbach (2000).

$$0.5 * E[(0.9 - I_{Heads})^2 | Red\ coin] + 0.5 * E[(0.1 - I_{Heads})^2 | Green\ coin]$$

$$= 0.5 * [(0.9) * (0.9 - 1)^2 + (0.1) * (0.9 - 0)^2] + 0.5 * [(0.1) * (0.1 - 1)^2 + (0.9) * (0.1 - 0)^2] = 0.09$$

Therefore, the relative improvement of B over A is $1 - \frac{0.09}{0.25} = 0.64$; that is, $CRPSS\ =\ 64\%$.
This skill score will be our main analytical tool for evaluating interest rate forecast performance and comparing results between different calibrations methods.

2.3 Probability vs. Frequency

We note that the evaluation criteria of the previous sections, that realized frequencies should agree in some sense with forecast probabilities, are based on a fundamentally "frequentist" view of probability, in that they presuppose the meaning of a probabilistic statement to be borne out in the frequencies of occurrence of events. While we do not seek to resolve here the centuries-old debate between this and other more "subjective" interpretations of probability, we should allow for the possibility that both the issuers and consumers of probabilistic forecasts may interpret them in other ways beyond simple statements about frequencies.

Practically, this may have to do with the availability of data for the entities involved in the forecast; for example, forecasts and validations for the Eurozone are limited to the period from 1999 to the present, so only a limited amount of validation data exists . Or, the forecasts may embed a forward-looking view about regulatory or policy regimes being different in the future from what they were in the past. It may also happen that forecasts involve probabilities for events that are extremely rare, historically unprecedented, or impossible to observe more than once, such as sovereign default, hyperinflation, economic collapse, etc., for which frequentist analyses are inherently difficult, bordering on inapplicable.

When dealing with extreme scenarios of this kind, prudence may recommend an allowance for "uncertainty" in the sense of Knight (1921) or Taleb (2007) to account for previously unanticipated possibilities with profound consequences. How much probability to assign to these scenarios, and how to assess their severity and magnitude, present difficult questions beyond the scope of this paper.

Even within a frequentist interpretation, it may be that such probabilistic statements are conditioned on so much information, such as current meteorological conditions or the present state of the global economy, that they resist validation over repeated trials, since the necessary conditions for validation never exactly repeat themselves. In the language of statistical mechanics, the conditioning variables constitute a "macrostate" of the world, summarized in key variables such as current interest rates and implied volatilities, while all other remaining possible variables (e.g., stock/commodity indices, currency exchange rates, GDP, unemployment figures, house prices, the dispositions of federal reserve members, etc.) – *and their full histories to date* – constitute the "microstate" that will actually determine the future evolution of the economy.

Exactly where to draw the border between macrostate and microstate can be difficult to decide, though. Obviously, forecasts will benefit from incorporating more available information; leaving less "up to chance" should make the forecasts more deterministic in principle, and assuming the resultant predictions are close to what is observed, the forecasting system will be rewarded accordingly. However, the danger of conditioning on too much information is that no common themes can emerge as relationships between variables, and so each forecast becomes an exercise unto itself; for example, if our forecast model for tomorrow's chance of rain depended on which day of the week it is currently, we would have no basis of connecting Monday's forecast with Tuesday's, or even necessarily considering them to be part of the same repeated phenomenon. So our data set becomes as fragmented as the complexity of our model demands, and we may not be left with enough data to draw any meaningful explanatory conclusions.

Generally speaking, if our goal is to produce models that express stable relationships we expect to persist through time, we should strive to make the models as simple as possible while incorporating all of the cogent information we think is relevant to the problem. However, it is also incumbent on us to document as clearly as possible those operating conditions we expect to be stable and borne out through frequencies vs. those events we assign probabilities to in order to indicate uncertainty.

3. Interest Rate Model Calibration

In this section we review the historical Moody's Analytics standard calibration methodology for real-world nominal interest rate models. We note here that the Extended 2-factor Black-Karasinski (E2FBK) model described below has since been superseded as the recommended model for interest rates by other, more sophisticated models (particularly a "displaced" version of E2FBK and a Libor Market Model variant with displacement) that are particularly well-suited to the current low-rate environments and have other desirable features. We confine our analysis here to the "old" model and associated calibration methods, but as more calibrations for the new models become available, we will seek to use this as a point of comparison between models.

These calibrations have been produced quarterly for the joint behavior of interest rates in 30 economies and subdivided into "Multi-year," used for long-term (10+ year) projections, and "1-year VaR,"[14] used primarily for 1-year projections for Value-at-Risk calculations and other short-term applications.

Nominal interest rates were simulated using the E2FBK model, which describes interest rate behavior via a mean-reverting process of two stochastic factors, correlated between economies,[15] and is initialized to match current market interest rates.[16] The resulting distributions of future interest rates are well approximated by a joint log-normal distribution.[17] In addition to the current yield curve, the other free parameters available each quarter in the 1-year VaR calibrations are:

- the mean-reversion rates and volatilities of the two stochastic factors, which are tuned to 1-year percentile targets based on market swaption implied volatility levels[18]
- the correlations between stochastic shocks of different economies, fixed at long-term levels
- the term premium (or "drift"), set to 0 for 1-year projections

We use the standard 1-year VaR calibrations and model 8 economies: USD, EUR, JPY, GBP, CAD, CHF, AUD, and HKD, for which we have available historical calibrations from June 2008 to December 2014. For each calibration quarter we generate a set of 10,000 scenarios using the ESG and fit a log-normal distribution to the one-quarter-ahead projected distribution of the 10-year nominal par yield. We use the fitted time-0 interest rate as validation against the previous quarter's projection.

Thus, our total data set consists of:

- 26*8 = 208 quarter-economies of projected rate distributions, with each quarter described by a joint log-normal distribution with a given mean vector and variance-covariance matrix
- 208 quarter-economies of the realized 10-year rate

We also consider two alternative calibration methodologies:

- a "Through-the-Cycle" calibration, with the same mean targets as the 1-year VaR calibrations above but with a variance-covariance structure consistent with long-term historical averages, as generated using the standard "Multi-year" calibration[19]
- a "Stationary" calibration, fit to the mean and variance-covariance of the realized (log-)interest rates themselves

The latter calibration is clearly unreasonable for the sake of forecasting, since the forecast distribution should not depend on the values being forecast, but will serve as a useful point of comparison. In particular, this helps to test against the hypothesis that the interest rates being considered are drawn from a stationary distribution without any significant time-dependent variability or serial correlation. This plays the role of the constant "climatological" forecast as in the discussion above.

[14] Also referred to as "conditional" calibrations. See Hibbert and Skrk (2008) for reference.

[15] See Morrison (2007) for details of the model.

[16] Either government rates, i.e., treasuries, or swap rates may be used. For the current analysis we use government calibrations.

[17] See Roseburgh and Morrison (2007) for a comparison of distributions.

[18] See Hibbert and Skrk (2008).

[19] See Tadrowski and Jessop (2012).

4. Forecast Analysis

For each forecast period we have a joint log-normal forecast distribution together with the realized interest rates for 8 economies: USD, EUR, JPY, GBP, CAD, CHF, AUD, and HKD. To begin with, we choose a permutation of the economies, for example:

1-USD, 2-EUR, 3-JPY, 4-GBP, 5-CAD, 6,-CHF, 7-AUD, 8-HKD

Then we transform the realized values $x_1, \ldots, x_8$ according to the multivariate Rosenblatt transform to obtain normalized values $z_1, \ldots, z_8$. Since the forecast distributions are log-normal, this is made simple by taking the Cholesky decomposition of the forecast variance-covariance matrix and multiplying by the vector of mean-centered values of the log interest rates. (See Appendix A for details).

Under the hypothesis that the forecast distributions are reliable, these z values are independent with uniform (0,1) distribution. We pool together results from all forecast periods and calculate the relevant one-sample Kolmogorov-Smirnov statistic against the uniform (0,1) distribution.

For each i, after conditioning on the values of $x_1, x_2, \ldots, x_{i-1}$ the forecast assumption is that x_i has a log-normal distribution with a given mean and variance. We calculate the Continuous Ranked Probability Score for the realized value against this distribution and average over all economies and forecast periods. For the sake of comparison, we also calculate the CRPS for the Through-the-Cycle and Stationary calibrations described above and compute skill scores (CRPSS) relative to each, with higher values indicating greater skill.

Finally, we repeat the above processes for all 8! = 40,320 possible permutations of the economies. Results are summarized in the table below:

	K-S Statistic: $\sqrt{N} * D_N$	CRPSS vs. Through-the-Cycle (best possible = 100%)	CRPSS vs. Stationary (best possible = 100%)
Median	1.3300	7.7%	3.8%
Min	0.5857	-4.5%	-25.0%
Max	2.3373	17.4%	18.9%

The critical values for the Kolmogorov-Statistic (N=208) are:

α	K-S Stat
0.2	1.07
0.1	1.22
0.05	1.36
0.01	1.63

So we see that the best result of the Kolmogorov-Smirnov test would not reject the null hypothesis that we are sampling from the same distributions at even the 20% level, whereas the worst result would reject at the 1% level (and indeed at the 0.1% level = 1.95). The median result would reject at the 10% level but not at the 5% level.

In the best permutation, the forecasts are seen to be skillful relative to both the Through-the-Cycle and Stationary calibrations, as in the median permutation. In the worst permutation, the forecasts have negative skill relative to both.

The differences in these permutation results are illustrated in the figures below, which show Q-Q plots of the pooled, transformed variables against their theoretical distribution in the best and worst permutation, measured by Kolmogorov-Smirnov statistic. For the sake of illustration, we show percentiles of $\Phi^{-1}(z)$ against those of the standard Normal distribution. Deviation below (above) the line $y = x$ indicates an upward (downward) bias in the forecasts; a line with slope greater than (less than) 1 indicates forecasts with too little (too much) volatility.

Figure 3: Q-Q Plot vs. N(0,1) Distribution [Best Permutation]

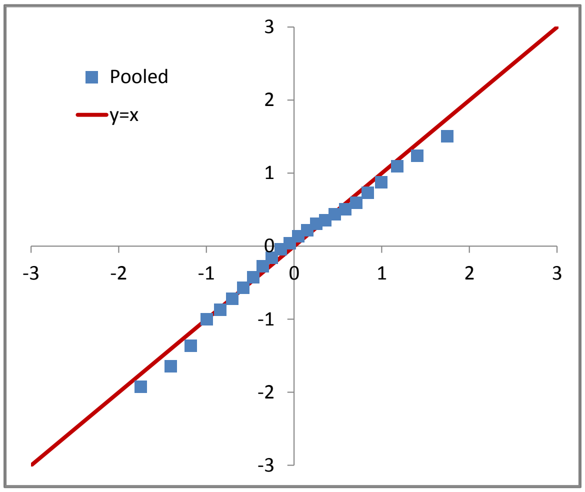
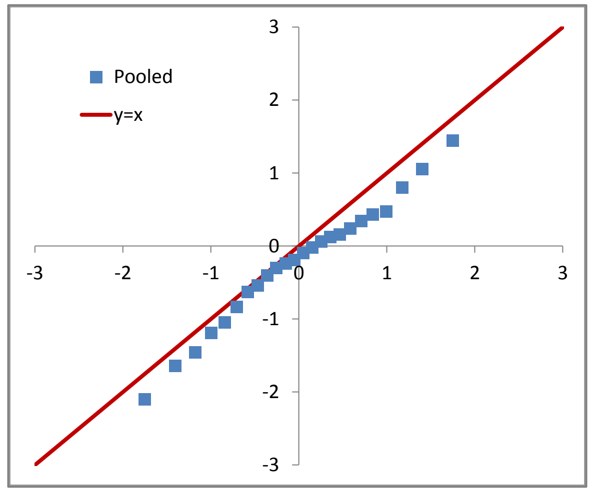


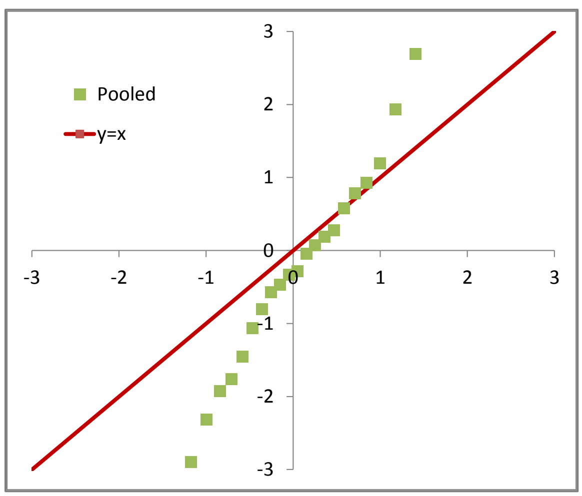
Figure 4: Q-Q Plot vs. N(0,1) Distribution [Worst Permutation]



These plots indicate that volatility levels are generally reliable, and some upward bias is seen in the forecasts, depending on the permutation order.

By comparison, the Through-the-Cycle calibration evidently forecast consistently too little volatility over the period:

Figure 5: Q-Q Plot vs. N(0,1) Distribution [TTC Calibration]



This pattern is seen in all permutations and is to be expected given that the (proportional) volatility assumptions in the multi-year calibrations reflect our view of volatility under average yield curve levels, while the historical data considered here corresponds to a period of relatively low interest-rates (and correspondingly high interest-rate volatility).

Examining some of the worst-performing permutations in detail reveals some joint relationships that did not validate well. In particular, many of the worst permutations had 1-USD and 2-CAD, which in isolation (that is, un-pooled), produced the Q-Q plots:

Figure 6: Q-Q Plot vs. N(0,1) Distribution [CAD]

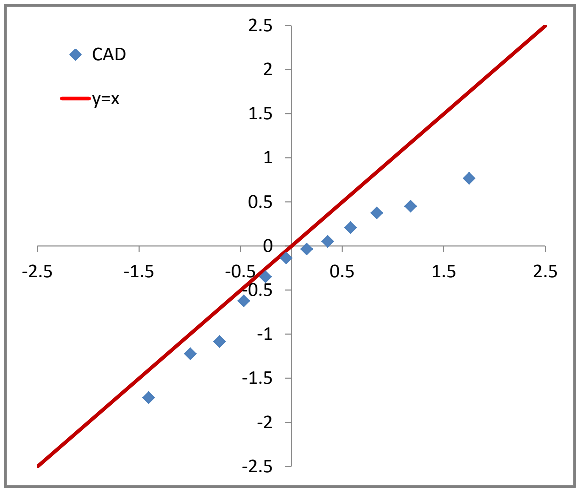
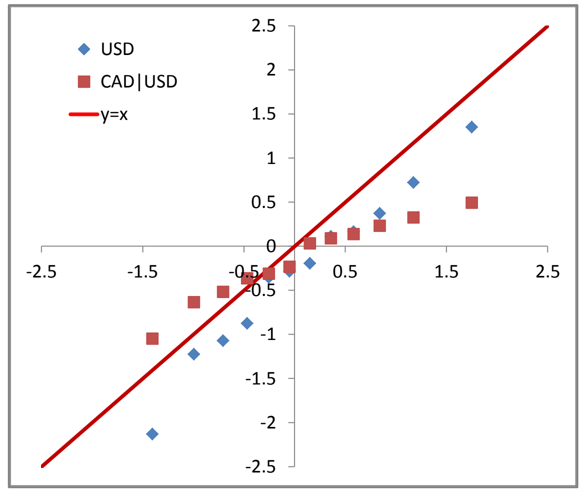


Figure 7: Q-Q Plot vs. N(0,1) Distribution [USD and CAD|USD]



We see that the CAD and USD forecasts had reasonably reliable volatility but were somewhat upwardly biased. The CAD forecasts, once conditioned on the realized USD rates, predicted too much volatility. This is consistent with a too-low correlation assumption between the two economies, meaning the conditional contribution of the USD interest rate did not account for a great enough portion of the total volatility of the CAD rate.

The joint dependency between economy pairs is dictated by the covariance matrix in the joint log-normal distribution, which results in turn from the correlation assumptions between stochastic shocks of the E2FBK model. We note that these shock correlations are not calibrated conditionally each quarter (in contrast to the volatility parameters) and are instead set equal to the Multi-year correlation assumptions. Furthermore, they are not economy-specific.

The resulting forecast distribution correlation matrices were therefore fairly stable from quarter to quarter and had a mostly uniform structure (with some variation due to the effect of rate levels):

Figure 8: Correlations in Changes in Log Rates [Forecast]

Correlation	USD	CAD	AUD	GBP	HKD	CHF	JPY	EUR
USD	100%	59%	61%	62%	60%	61%	45%	61%
CAD	59%	100%	63%	61%	60%	61%	43%	60%
AUD	61%	63%	100%	64%	62%	64%	41%	63%
GBP	62%	61%	64%	100%	61%	63%	44%	61%
HKD	60%	60%	62%	61%	100%	61%	47%	60%
CHF	61%	61%	64%	63%	61%	100%	44%	62%
JPY	45%	43%	41%	44%	47%	44%	100%	47%
EUR	61%	60%	63%	61%	60%	62%	47%	100%

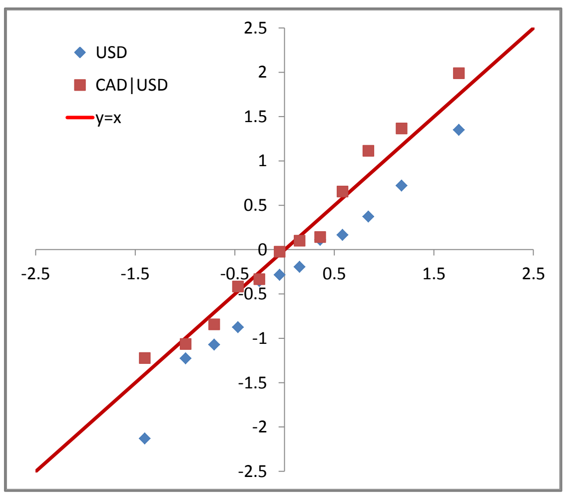
By contrast, calculating summary correlations in changes in (log-) interest rates over the period indicates areas in which realized correlations were much higher, with a clearly distinguishable grouping of economies (USD/CAD/AUD/GBP):

Figure 9: Correlations in Changes in Log Rates [Realized]

Correlation	USD	CAD	AUD	GBP	HKD	CHF	JPY	EUR
USD	100%	96%	84%	90%	75%	77%	58%	73%
CAD	96%	100%	81%	93%	69%	79%	64%	74%
AUD	84%	81%	100%	83%	75%	75%	55%	67%
GBP	90%	93%	83%	100%	65%	82%	67%	76%
HKD	75%	69%	75%	65%	100%	66%	35%	64%
CHF	77%	79%	75%	82%	66%	100%	50%	80%
JPY	58%	64%	55%	67%	35%	50%	100%	67%
EUR	73%	74%	67%	76%	64%	80%	67%	100%

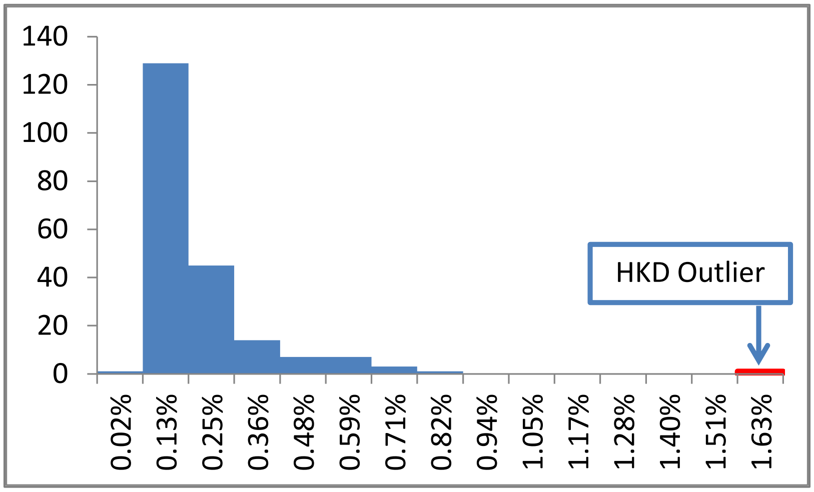
In particular, the realized value of the USD/CAD correlation was **96%**, compared to a constant forecast correlation assumption of **59%**. Changing this forecast assumption to be in agreement with its realized level appears to improve the CAD conditional performance dramatically:

Figure 10: Q-Q Plot vs. N(0,1) Distribution [USD and CAD|USD; Higher Correlation Assumption]



One clear outlier emerged, regardless of permutation order: the HKD 10-year rate for December 2008 fell to **1.18%** from its previous value at September 2008 of **2.87%**. According to the forecast distribution, this represented a -**10** standard deviation change in the log of the rate; that is, a movement this extreme was assigned a probability $< 10^{-24}$. Notably, this outlier was not detected by the Kolmogorov-Smirnov test, since it accounts for only a very small difference (< 0.005) between theoretical and empirical c.d.f. at this value, but *was* penalized heavily in the CRPS calculation due to the large absolute deviation in forecast value. The histogram of CRPS values over the observations clearly shows the outlier:

Figure 11: Histogram of Continuous Ranked Probability Scores [N=208]



We can also perform similar analysis for the one-**year**-ahead forecasts, in addition to the one-quarter-ahead, but due to the very limited data set [N=48] it becomes difficult to draw any meaningful conclusions. For example, Q-Q Plots for the pooled 1-year forecasts (grouped according to calibration quarter, that is, Q1-Q1, Q2-Q2, etc.) showed some evidence of upward bias in the forecasts:

Figure 12: Q-Q Plot vs. N(0,1) Distribution for Annual Forecasts [Best Permutation]

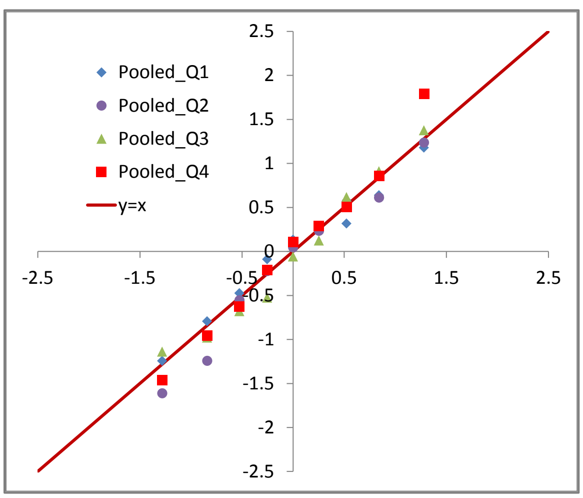
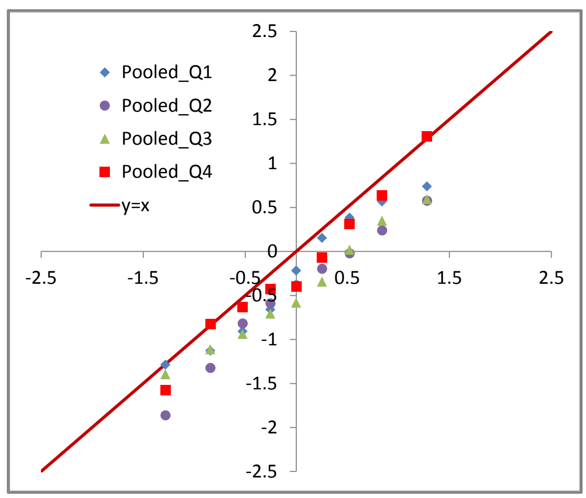


Figure 13:: Q-Q Plot vs. N(0,1) Distribution for Annual Forecasts [Worst Permutation]



These results are highly dependent on permutation order and subject to a large amount of statistical noise.

Kolmogorov-Smirnov tests for the annual forecasts either reject the null hypothesis at the 0.1% level or fail to reject at the 20% level depending on the permutation.

5. Conclusions

We have presented an analytical framework to evaluate multivariate probabilistic forecasts, such as economic variable forecasts for multiple economies. This extends the back-testing methods used previously in the following ways:

- We analyze the forecasts for all variables simultaneously and include statistical tests that are sensitive to the full joint distribution across economies, instead of treating each economy in isolation.
- We separate the forecast "reliability," in the sense of agreement between forecast distributions vs. realized frequencies, and the forecast "resolution," in the sense of the model's ability to identify conditions in which realized frequencies should be expected to vary.
- We express these measures in relative skill scores that allow for quantitative assessment of performance against other models and calibrations.

The results of our analysis suggest that the Moody's quarterly interest rate forecast distributions are generally reliable for one-quarter-ahead predictions. Volatility levels appear to be well calibrated, but some upward bias is present in the results, suggesting the mean targets may be consistently higher than realized rates. A possible refinement to address this would be setting a negative drift term in the E2FBK interest rate model, corresponding to a short-term positive term premium on bonds. Skill scores indicate the conditional forecasts are able to resolve periods of high and low volatility, and these conditional forecasts generally outperform Through-the-Cycle and Stationary calibrations, although this depends on the order in which economies are considered. This suggests any improvement should likely focus on the inter-economy relationships.

Analyzing results further for some particular economy pairs, most notably USD/CAD, suggests joint distributional assumptions that disagree somewhat with realized events over the period of time, albeit in a relatively small sample of only 26 quarters. Forecast correlations are generally substantially lower than realized. If these joint relationships are deemed to be predictable, it would significantly improve the calibration performance to calibrate the correlations in model shocks conditionally and in an economy-specific manner.

An extreme outlier in HKD suggests the realized change from September to December 2008 was not adequately captured by the stochastic model in that economy. Further analysis could address which model assumptions were violated in this instance, and whether this kind of uncertain phenomenon could be incorporated into the model in a different way.

Due to the paucity of available data, we have pooled together model forecasts from multiple economies and over multiple time periods, using standard transformation techniques that may obscure some biases that are local to particular forecasts. As more calibrations become available (either in future or in historical back-filling), this analysis could be refined further by focusing on particular economies or interest rate environments. Analyzing swap rates, inflation-indexed yields, or other points on the yield curve (e.g., the short rate) could also be informative.

Finally, as models are improved or enhanced to incorporate different dynamics (e.g., forward-rate models or models with "displacement" terms[20]), the above analysis could be expanded to validate these models and compare performance of the different models against each other. Naturally, different models will embed different distributional assumptions, but the generic methods we have presented of normalizing and scoring forecast performance would allow them to be evaluated on an equal footing.

[20] For example, Hibbert and Jessop (2012) consider a displaced Libor Market Model with stochastic volatility, and Skrzyszowski (2015) considers a displaced version of the E2FBK model described here.

A. Appendix – Derivation of Conditional Distributions for Jointly Log-Normal Variables

We assume a joint log-normal distribution for interest rates $X_1, \dots, X_N$ along with their realized values $x_1, \dots, x_N$. Let $Y_i = \log(X_i)$ and $y_i = \log(x_i)$ for all i, and assume the joint distribution of the Y_i is parameterized by the mean vector

$$\mu = (\mu_1, \dots, \mu_N)^T$$

and the variance-covariance matrix

$$\Sigma = \left(Cov(Y_i, Y_j)\right)_{1 \le i,j \le N}$$

We seek the conditional distribution of each Y_{k+1} conditional on $Y_1 = y_1, \dots, Y_k = y_k$. For notational simplicity, write $(y-\mu)_k$ for the (column) vector of mean-centered values

$$(y-\mu)_k = \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \\ \dots \\ y_k - \mu_k \end{bmatrix}$$

It suffices to compute the conditional mean and variance of Y_{k+1}. To that end, consider the Cholesky decomposition of Σ:

$$\Sigma = LL^T$$

for a lower triangular matrix L.

In particular, it follows that $L^{-1}(Y-\mu)$ has the standard multivariate normal distribution, $N(0, I_N)$.

Letting

$$L^{(k)} := \left(L_{ij}\right)_{1 \le i,j \le k}$$

be the top-left kxk sub-matrix of L, and

$$L_{k+1} := \left[L_{k+1,1}\; L_{k+1,2} \dots L_{k+1,k}\right]$$

be the row vector of the first k entries of the $(k+1)^{th}$ row, it follows from the fact that L is lower triangular that we can write

$$Y_{k+1} = \mu_{k+1} + L_{k+1}[L^{(k)}]^{-1}(y-\mu)_k + L_{k+1,k+1}Z_{k+1}$$

for some variable Z_{k+1} with the standard $N(0,1)$ distribution, from which we have the conditional mean of Y_{k+1}

$$E[Y_{k+1} | Y_1 = y_1, \dots, Y_k = y_k] = \mu_{k+1} + L_{k+1}[L^{(k)}]^{-1}(y-\mu)_k$$

and the conditional variance

$$Var[Y_{k+1} | Y_1 = y_1, \dots, Y_k = y_k] = L_{k+1,k+1}^2$$

References

Brier, Glenn W. "Verification of Forecasts Expressed In Terms of Probability." *Monthly Weather Review*, 1950, 1-3.

Conn, Gavin, Harry Hibbert, Sandy Sharp, and Craig Turnbull. "Validation of Risk Factor Modelling in 1-year VaR Capital Assessments." *Moody's Analytics Research*, 2013. http://www.barrhibb.com/research_and_insights/article/validation_ofr_risk_factor_modelling_in_1-year_var_capital_assessment.

Crossen, Christopher, Shisheng Qu, and Xu Zhang. "Validating the Public EDF Model for North American Corporate Firms." *Moody's Analytics Methodology*, 2011.

Diebold, Francis X., Jinyong Hahn, and Anthony S. Tay. "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange." *Review of Economics and Statistics*, 1999, 661-73.

Dwyer, Douglas and Irina Korablev. "Power and Level Validation of Moody's KMV EDF™ Credit Measures in North America, Europe, and Asia." *Moody's KMV*, 2007.

El Cherif, Frederic. "Real-world Equity Calibration: A Comparison of Realised and Expected Volatility." 2010. http://www.barrhibb.com/knowledge_base/article/real-world_equity_calibration_a_comparison_of_realised_and_expected_vol_2/.

Epstein, Edward S. "A Scoring System for Probability Forecasts of Ranked Categories." *Journal of Applied Meteorology*, 1969, 985-87.

Hersbach, Hans. "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems." *Weather and Forecasting*, 2000, 559-70.

Hibbert, Harry, and Nick Jessop. "Real-world One-year Nominal Interest Rates: LMM Plus Calibration." 2012. http://www.barrhibb.com/knowledge_base/article/real-world_one-year_nominal_interest_rates_lmm_plus_calibration/.

Hibbert, John, and Martin Skrk. "Real-World Interest Rate Distribution: End 2007 - 1-year Ahead Interest Rate Tails: Initial Analysis and Discussion." 2008. http://www.barrhibb.com/knowledge_base/article/1_year_ahead_interest_rate_tails/.

Hu, Zhenya, Amnon Levy, and Jing Zhang, "Economic Capital Model Validation: A Comparative Study." *Moody's Analytics Methodology*, 2012.

Jolliffe, Ian T., and David B. Stephenson, eds. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Chichester, West Sussex: J. Wiley, 2012.

Knight, Frank H. *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin, 1921.

Matheson, James E., and Robert L. Winkler. "Scoring Rules for Continuous Probability Distributions." *Management Science*, 1976, 1087-096.

Morrison, Steven. "Implementation of the Extended 2-Factor Black-Karasinski Model." 2007. http://www.barrhibb.com/knowledge_base/article/implementation_of_the_extended_2_factor_black_karasinski_model/.

Murphy, Allan H. "A New Vector Partition of the Probability Score." *Journal of Applied Meteorology*, 1973, 595-600.

Murphy, Allan H., and Edward S. Epstein. "Verification of Probabilistic Predictions: A Brief Review." *Journal of Applied Meteorology*, 1967, 748-55.

Murphy, Allan H., and Robert L. Winkler. "Diagnostic Verification of Probability Forecasts." *International Journal of Forecasting*, 1992, 435-55.

Roseburgh, David, and Steven Morrison. "Approximating Interest Rate Distributions Within 2FBK." 2007. http://www.barrhibb.com/knowledge_base/article/approximating_interest_rate_distributions_within_2fbk/.

Rosenblatt, Murray. "Remarks on a Multivariate Transformation." *Ann. Math. Statist.* 23, no. 3 (1952): 470-72.

Skrzyszowski, Karol. "Extended Displaced 2-factor Black-Karasinski Model: Govt Calibrations and Validations." 2015. http://www.barrhibb.com/knowledge_base/article/extended_displaced_2-factor_black-karasinski_model_govt_calibrations_and_va/.

Tadrowski, Laura, and Nick Jessop. "Multi Year Real-World Interest Rates: Updating E2FBK, 2F Vasicek and Inflation Plus Calibrations." 2012.

http://www.barrhibb.com/knowledge_base/article/multi_year_real_world_interest_rates_updating_e2fbk_2f_vasicek_and_inflatio/
.

Taleb, Nassim Nicholas. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House, 2007.

© 2015 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "MOODY'S"). All rights reserved.

CREDIT RATINGS ISSUED BY MOODY'S INVESTORS SERVICE, INC. ("MIS") AND ITS AFFILIATES ARE MOODY'S CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES, AND CREDIT RATINGS AND RESEARCH PUBLICATIONS PUBLISHED BY MOODY'S ("MOODY'S PUBLICATIONS") MAY INCLUDE MOODY'S CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES. MOODY'S DEFINES CREDIT RISK AS THE RISK THAT AN ENTITY MAY NOT MEET ITS CONTRACTUAL, FINANCIAL OBLIGATIONS AS THEY COME DUE AND ANY ESTIMATED FINANCIAL LOSS IN THE EVENT OF DEFAULT. CREDIT RATINGS DO NOT ADDRESS ANY OTHER RISK, INCLUDING BUT NOT LIMITED TO: LIQUIDITY RISK, MARKET VALUE RISK, OR PRICE VOLATILITY. CREDIT RATINGS AND MOODY'S OPINIONS INCLUDED IN MOODY'S PUBLICATIONS ARE NOT STATEMENTS OF CURRENT OR HISTORICAL FACT. MOODY'S PUBLICATIONS MAY ALSO INCLUDE QUANTITATIVE MODEL-BASED ESTIMATES OF CREDIT RISK AND RELATED OPINIONS OR COMMENTARY PUBLISHED BY MOODY'S ANALYTICS, INC. CREDIT RATINGS AND MOODY'S PUBLICATIONS DO NOT CONSTITUTE OR PROVIDE INVESTMENT OR FINANCIAL ADVICE, AND CREDIT RATINGS AND MOODY'S PUBLICATIONS ARE NOT AND DO NOT PROVIDE RECOMMENDATIONS TO PURCHASE, SELL, OR HOLD PARTICULAR SECURITIES. NEITHER CREDIT RATINGS NOR MOODY'S PUBLICATIONS COMMENT ON THE SUITABILITY OF AN INVESTMENT FOR ANY PARTICULAR INVESTOR. MOODY'S ISSUES ITS CREDIT RATINGS AND PUBLISHES MOODY'S PUBLICATIONS WITH THE EXPECTATION AND UNDERSTANDING THAT EACH INVESTOR WILL, WITH DUE CARE, MAKE ITS OWN STUDY AND EVALUATION OF EACH SECURITY THAT IS UNDER CONSIDERATION FOR PURCHASE, HOLDING, OR SALE.

MOODY'S CREDIT RATINGS AND MOODY'S PUBLICATIONS ARE NOT INTENDED FOR USE BY RETAIL INVESTORS AND IT WOULD BE RECKLESS FOR RETAIL INVESTORS TO CONSIDER MOODY'S CREDIT RATINGS OR MOODY'S PUBLICATIONS IN MAKING ANY INVESTMENT DECISION. IF IN DOUBT YOU SHOULD CONTACT YOUR FINANCIAL OR OTHER PROFESSIONAL ADVISER.

ALL INFORMATION CONTAINED HEREIN IS PROTECTED BY LAW, INCLUDING BUT NOT LIMITED TO, COPYRIGHT LAW, AND NONE OF SUCH INFORMATION MAY BE COPIED OR OTHERWISE REPRODUCED, REPACKAGED, FURTHER TRANSMITTED, TRANSFERRED, DISSEMINATED, REDISTRIBUTED OR RESOLD, OR STORED FOR SUBSEQUENT USE FOR ANY SUCH PURPOSE, IN WHOLE OR IN PART, IN ANY FORM OR MANNER OR BY ANY MEANS WHATSOEVER, BY ANY PERSON WITHOUT MOODY'S PRIOR WRITTEN CONSENT.

All information contained herein is obtained by MOODY'S from sources believed by it to be accurate and reliable. Because of the possibility of human or mechanical error as well as other factors, however, all information contained herein is provided "AS IS" without warranty of any kind. MOODY'S adopts all necessary measures so that the information it uses in assigning a credit rating is of sufficient quality and from sources MOODY'S considers to be reliable including, when appropriate, independent third-party sources. However, MOODY'S is not an auditor and cannot in every instance independently verify or validate information received in the rating process or in preparing the Moody's Publications.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability to any person or entity for any indirect, special, consequential, or incidental losses or damages whatsoever arising from or in connection with the information contained herein or the use of or inability to use any such information, even if MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers is advised in advance of the possibility of such losses or damages, including but not limited to: (a) any loss of present or prospective profits or (b) any loss or damage arising where the relevant financial instrument is not the subject of a particular credit rating assigned by MOODY'S.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability for any direct or compensatory losses or damages caused to any person or entity, including but not limited to by any negligence (but excluding fraud, willful misconduct or any other type of liability that, for the avoidance of doubt, by law cannot be excluded) on the part of, or any contingency within or beyond the control of, MOODY'S or any of its directors, officers, employees, agents, representatives,licensors or suppliers, arising from or in connection with the information contained herein or the use of or inability to use any such information.

NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE ACCURACY, TIMELINESS, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OF ANY SUCH RATING OR OTHER OPINION OR INFORMATION IS GIVEN OR MADE BY MOODY'S IN ANY FORM OR MANNER WHATSOEVER.

MIS, a wholly-owned credit rating agency subsidiary of Moody's Corporation ("MCO"), hereby discloses that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by MIS have, prior to assignment of any rating, agreed to pay to MIS for appraisal and rating services rendered by it fees ranging from $1,500 to approximately $2,500,000. MCO and MIS also maintain policies and procedures to address the independence of MIS's ratings and rating processes. Information regarding certain affiliations that may exist between directors of MCO and rated entities, and between entities who hold ratings from MIS and have also publicly reported to the SEC an ownership interest in MCO of more than 5%, is posted annually at www.moodys.com under the heading "Shareholder Relations — Corporate Governance — Director and Shareholder Affiliation Policy."

For Australia only: Any publication into Australia of this document is pursuant to the Australian Financial Services License of MOODY'S affiliate, Moody's Investors Service Pty Limited ABN 61 003 399 657AFSL 336969 and/or Moody's Analytics Australia Pty Ltd ABN 94 105 136 972 AFSL 383569 (as applicable). This document is intended to be provided only to "wholesale clients" within the meaning of section 761G of the Corporations Act 2001. By continuing to access this document from within Australia, you represent to MOODY'S that youare, or are accessing the document as a representative of, a "wholesale client" and that neither you nor the entity you represent will directly or indirectly disseminate this document or its contents to "retail clients" within the meaning of section 761G of the Corporations Act 2001. MOODY'S credit rating is an opinion as to the creditworthiness of a debt obligation of the issuer, not on the equity securities of the issuer or any form of security that is available to retail clients. It would be dangerous for "retail clients" to make any investment decision based on MOODY'S credit rating. If in doubt you should contact your financial or other professional adviser.