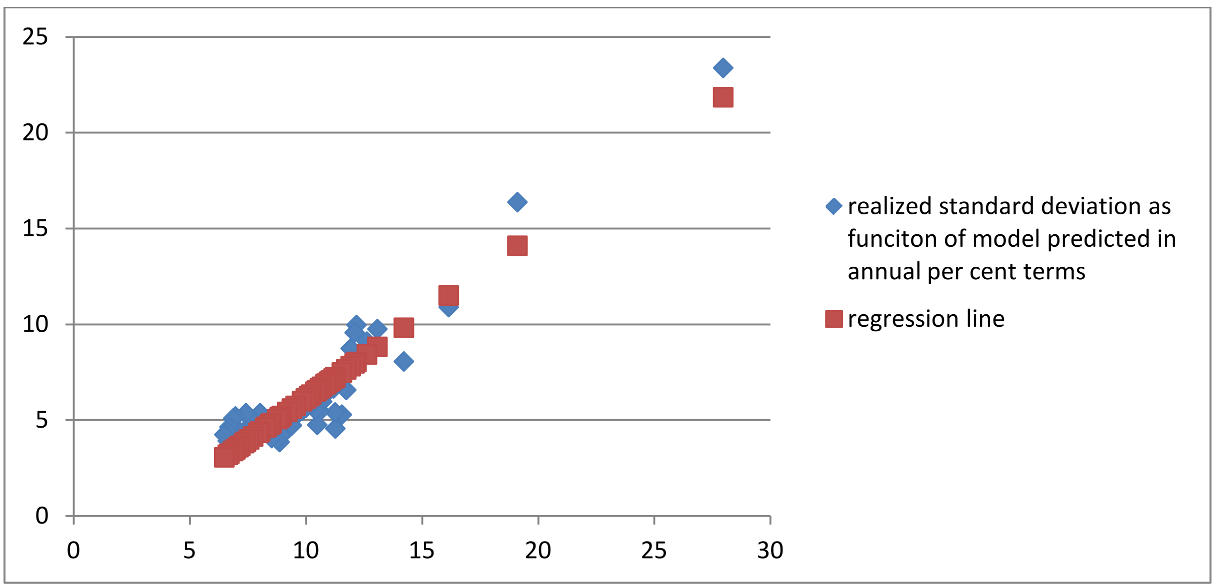
Bond Modeling in the Everything Everywhere Risk Model: Test Results

From Emilian Belev, CFA

It is our pleasure to present the EE model's test results. They are a product from our extensive testing on a universe of the 27 thousand most actively traded bonds around the world, as well as portfolios thereof, in which we benchmarked their predicted and realized risk performance in the first half of this year, specifically February to June 2011. The realized (or observed) risk performance was based on bond prices sourced from one of the largest and most respected commercial bond data providers in the industry.

There were four tests performed. In the first test we took 50 random portfolios of 100 bonds each and compared model predicted risk with realized volatility over the period. Each portfolio was picked so that it is biased towards a particular bond duration band. Given the dominant role of interest rate sensitivity on bond risk this should be considered as the most appropriate broad test for the model. Here is what we found:



The predicted vs. realized variance correlation is **0.97**. The R-squared of the same relationship is **0.94**.

The slope regression coefficient in the graph above is 0.87 and it has a very high statistical significance with a t-statistic of 18.6[1].

As an issue of high importance, in the same regression we also found that there is a statistically significant intercept (t-statistic = -5.13) of -2.6%. **It can be rashly interpreted as an upward bias in the model risk prediction, but we want to assert that this is not the case.** The rationale for our assertion is as follows.

[1] Note that in the version of the EE model prior to the firm sustainability credit was introduced, when the same type of regression was performed, the slope coefficient was 0.74 with a t-statistic of 16 and an R-squared of 0.84 - all values a somewhat lower than the current ones. Hence, the firm sustainability credit approach introduced a moderate but marked improvement in the ability of the model to predict risk accurately.

There are three facts of importance to our argument. First, each of the random portfolios on which the test was performed was biased towards a different duration level. Second, the amount of the difference between predicted and realized risk across portfolios is approximately constant. Third, the predicted risk for all portfolios consists of exclusively factor risk, with negligible specific risk, the reason being that each portfolio is comprised of hundreds of bonds from various issuers and is fully diversified.

We would first argue that most commercially accessible bond pricing processes are prone to have one disadvantage when it comes to estimating volatility– an element of smoothing. Bonds, as a vast majority, are not traded continuously as stocks, and the gaps in traded prices need to be filled. The data provider's business is to furnish prices on a daily basis, even in the absence of trades for a particular bond. In the absence of a model of risk drivers, it is hard to come up with a rigorous way to fill in the gaps. The tool at disposal generally lean towards interpolation. This is understandable - the pricing source prime purpose is to come up with a price, and in the common sense thing to do is to look at the neighboring values. And this does not undermine the practical accuracy of the daily price value. However, when one utilizes such prices to estimate bond return volatility, one in essence is compounding this smoothing effect to an ostensible level. Smoothing, of course, means less volatility of the reported time series.

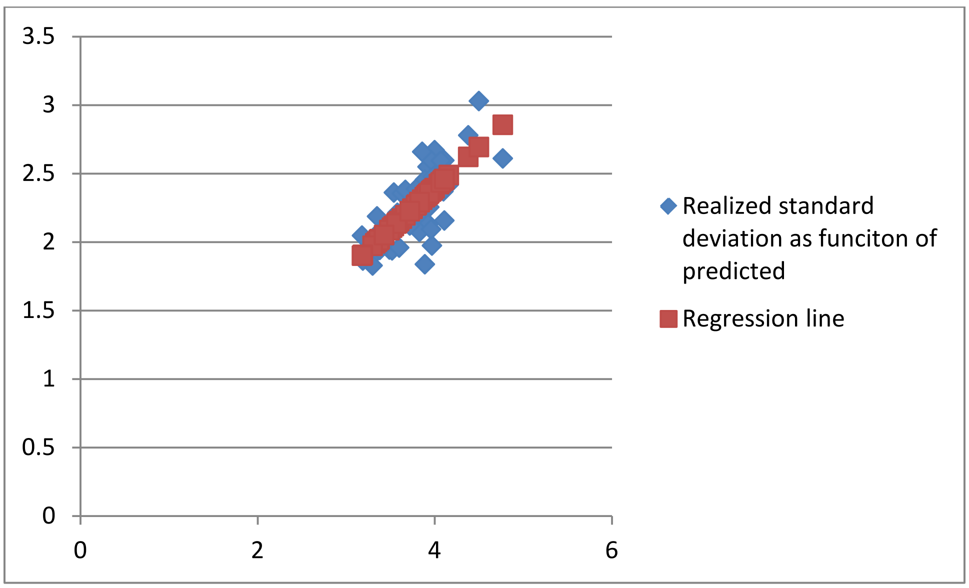
The key to understanding this issue is that we, at Northfield, are not trying to explain the pricing source's observed return volatility, but actually both we and the pricing source, in different ways are trying to estimate an unobservable variable – the "what if" continuous bond return volatility. Many bonds are not traded every day, but we would like to know what it will be if they did – i.e. what is the risk inherent if we try to sell this bond today. The only way we differ is that Northfield and the pricing source try to accomplish this estimation in different ways – we create a risk model, and the pricing source hand collects values, interpolates where needed, and enters values in a database.

The immediately good news for both us and the pricing source is the perfect correlation between our estimate and theirs. Given the complete independence of the sources and methods of data derivation, the maximum likelihood inferences is that both organizations are doing a good job.

We take this line of analysis further. If the difference between our estimate of standard deviation and pricing source's estimate was risk "bias" contributed by Northfield, then it will necessarily be factor risk error, because we report no specific risk for those test portfolios, and so it will likely be a different difference for the different portfolios as the different test portfolios have different factor exposures (notably durations). But the difference is [almost] constant across portfolios which contradicts the hypothesis that this is factor risk. And since Northfield estimate has nothing but factor risk this also contradict the hypothesis that our estimate is the one contributing to the difference. Since we call in broad concepts from the financial world (EE factors), which in essence are easier to accept as epitomizing the forces driving risk, than the demonstrably idiosyncratic nature of a data collecting methodology of the pricing source, it is sound statistical logic to state that our estimate is more likely to be right. In that sense we would assert that our estimates are right and the so called "realized" volatility has a downward bias.

Another contributor with equal importance to this uniform difference is the survivorship bias. Bonds that are priced over the test period are necessarily bonds in good standing over the same period, i.e. they have not defaulted. The ex-ante prediction of risk however, assumes that default for those bonds is possible. In that sense the "liquid bond" group is a biased sample towards "survivors" of the odds. This creates a divergence between the unbiased view of the risk model, and the biased sample. Moreover, the liquid bond group is biased towards higher credits, which potentially cannot improve much but potentially but can go to default. This makes the sample even more biased, enlarging the difference between model predicted and realized bond and portfolio risk.

As a continuation of our test sequence, a variation of the 50 portfolio test was also performed, this time with completely random broad portfolios. Those portfolios being big samples of a population have very similar characteristics, and the quality of the results are dominated by the sameness. Yet, even in this setting the high correlation between realized and predicted risk of 0.73 and regression explanation (slope of 0.6 with t-statistic of 7.5), prove the ability of the model to discern risk within a narrow range of risk values.

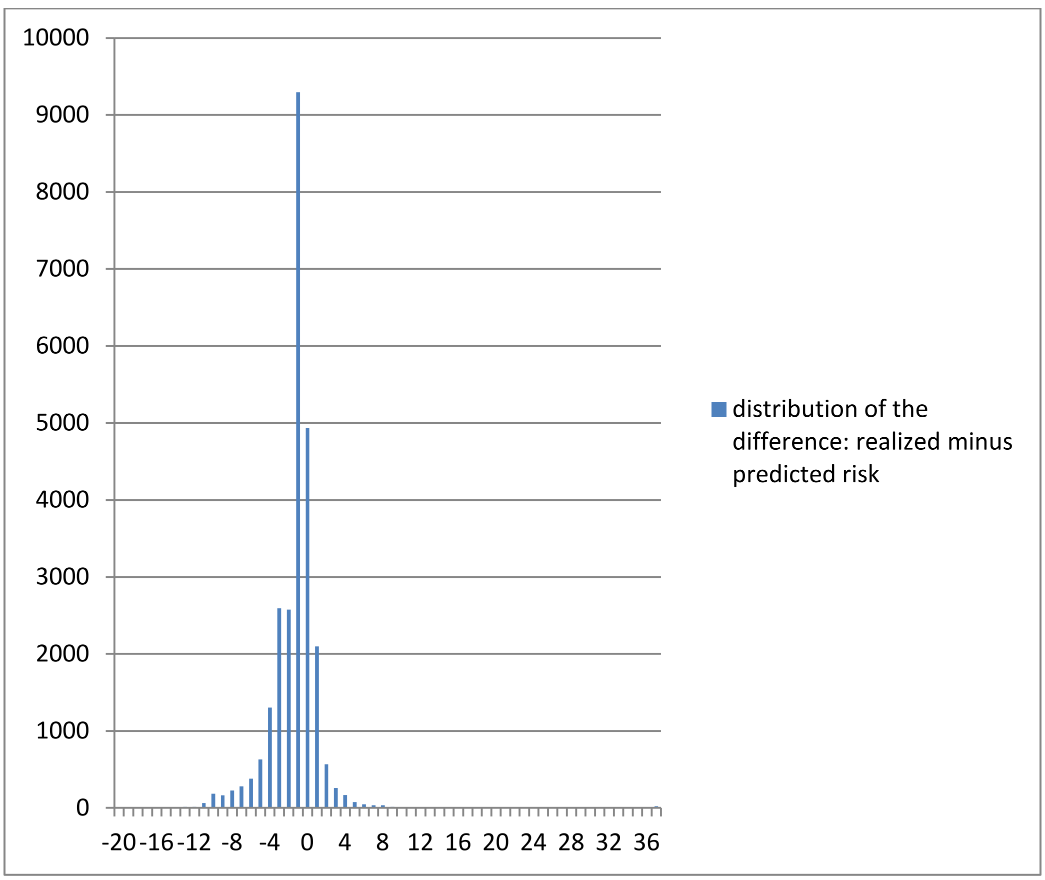


Another test was performed to look at credit risk in isolation . The test takes the credit buckets from the new model, where in essence each synthetic record is an equal-weighted portfolio of the bonds in that bucket, and aggregates the buckets per credit rating over sectors and region. The resulting "total" risk was only credit induced risk – currency and treasury factor exposures were removed. Effectively, the test is based on a pool of close to a hundred thousand bonds.

As can be intuitively expected, the volatility goes clearly from higher to lower as we go from lower to higher credit quality, as defined by the rating agencies. The only marked exception is the lowest rating C - it has a lower credit volatility than the next rating up. This make sense, given that such debtors are practically in default and for them the certainty of the debt dues demanded dwarf any business level uncertainty from the non-performing economic assets.

Rating	Credit Volatility - Annual stdev %
AAA	0.260628455
AA	0.420209
A	1.393854161
BBB	2.060612143
BB	3.9441636
B	5.623509565
CCC	4.817308667
CC	6.447296667
C	4.323928

The last test was at the individual bond level. As we did initially for random portfolios, we compared the realized and predicted risk of individual bonds. The vast majority (60%) of bonds came in within 1% realized historic standard deviation **under** predicted. About 90% of the bonds came within 3% realized historic standard deviation **under** predicted. We have to remember the smoothing component of "observed" bond volatility revealed by the portfolio level analysis, and the survivorship bias of liquid bonds mentioned earlier. Undoubtedly, both play a role here and we should necessarily explain those constant level differences between the risk model results and the results derived from the data from the pricing source with these two phenomena.



Conclusion

All the result presented here are even more impressive in view of the fact that the model that produced them builds on the very basic economics of the instruments, not from a statistical technique that starts off with the modeled data (regressions, PCA, etc.) Statistical techniques often guarantee, at least at the individual instrument level, that the breakdown in risk sources will sum up to the modeled variable. On the other hand, working the modeled variable up from the very fundamental characteristics of the financial instrument offers no such guarantee, unless such methodology is the right one. The excellent agreement between our model results and those of the pricing source's are a credit to our sustained endeavor for harmonizing statistical methodology and economic intuition.