RiskMetrics Group

Research Monthly

Good, bad or useful?

June 2008

Christopher C. Finger
chris.finger@riskmetrics.com

All models are wrong; some models are useful.

At risk management conferences, an invocation of George Box's dictum above is usually a precursor to an apologetic defense of a model, with the general theme of the talk being "I know you've heard a lot of terrible things, but trust me, if you look at it right, this model still works."

But this is not a model apology. We are certainly not the first to criticize the standard model for portfolio credit derivatives. Our point here, though, is as much to criticize the criticizers, and to suggest what should be the criteria on which the current standard model, as well as pretenders to the crown of the new standard, are judged. We are not usually fond of urging anyone to throw out the models, but it is important that if we do, we do it for the right reasons. If all models are indeed wrong, then this comes down to how we decide whether a model is useful.

To start, we should specify how the model is used. In the space of portfolio credit derivatives, specifically credit index tranches, as a prerequisite, we would first like a model to explain, or calibrate to, a set of standard products, in both normal and stressed market environments. In many cases, this is where the evaluation of a model begins and ends, which is a mistake. More than just matching prices, the model should prove useful for one of three applications.

First, we use the model to extrapolate prices, that is, to value custom derivatives that are similar to the standard tranches. Second, we hedge the index tranches with the underlying credits. For a market maker, the ability to hedge is crucial to providing liquidity in the standard tranches; for a speculator, hedges allow for positioning on the relative value of the tranches while maintaining a neutral position on the underlying credits. Finally, for risk management, we aggregate our exposures to individual credits across simple and complex positions, and identify, with tranche positions, how much risk is truly attributable to underlying credits and how much to the idiosyncracies of the derivative.

The first of these applications is ill suited to empirical validation, since the products we need to value have no observable price. Certainly, though, any other empirical backing would help validate the model for this application. The second and third applications are better suited to our needs. Because we consider standard tranches with standard underlying underlyings, we can observe prices on everything and evaluate whether the hedges from a model are in fact useful. And we can investigate whether the risks that our model deems idiosyncratic are in fact uncorrelated with the underlying factors. These will be our two yardsticks for evaluating models, and what we propose as the proving ground for any new contender.

In this corner . . .

We consider standard credit derivatives products, beginning with credit default swaps (CDS) referencing a single issuer, or name. Next, we consider the standard credit derivative indices, each a single contract referencing a portfolio of single-name CDS. We restrict our focus to the CDX North American Investment Grade index, which referencing 125 underlying constituent names. The index trades as a contract in its own right, and its price typically includes a small basis relative to the price implied by its constituents.[1] Last, we consider the standard tranches, derivatives referencing specific index losses. The most junior standard tranche on the CDX references losses up to 3%, and the most senior standard tranche references losses from 15% to 30% of the index portfolio.

So what is the standard model for credit index tranches? First, we derive default probabilities for individual names from the market for simple credit derivatives, either CDS or credit indices. Up to interpolation (that is, how do we arrive at a four-year default probability from market quotes for only three- and five-year maturities), the simple credit derivatives define everything we need to know about what happens to individual names in isolation.

The individual names are non-controversial enough, leaving us with how to express portfolio effects. The standard is to link the individual names through a one-factor Gaussian copula (OFGC). The Gaussian copula boils down to assuming that the time to default of each name is driven by some unseen normally distributed random variable; the one factor refers to the further assumption that the normal variables for the names are linked by a common exposure to a single common source of risk, and that therefore each pair of names shares a single value for their correlation. This second assumption is particularly enticing, in that it reduces the entire description of the correlation structure to a single number.

31 flavors

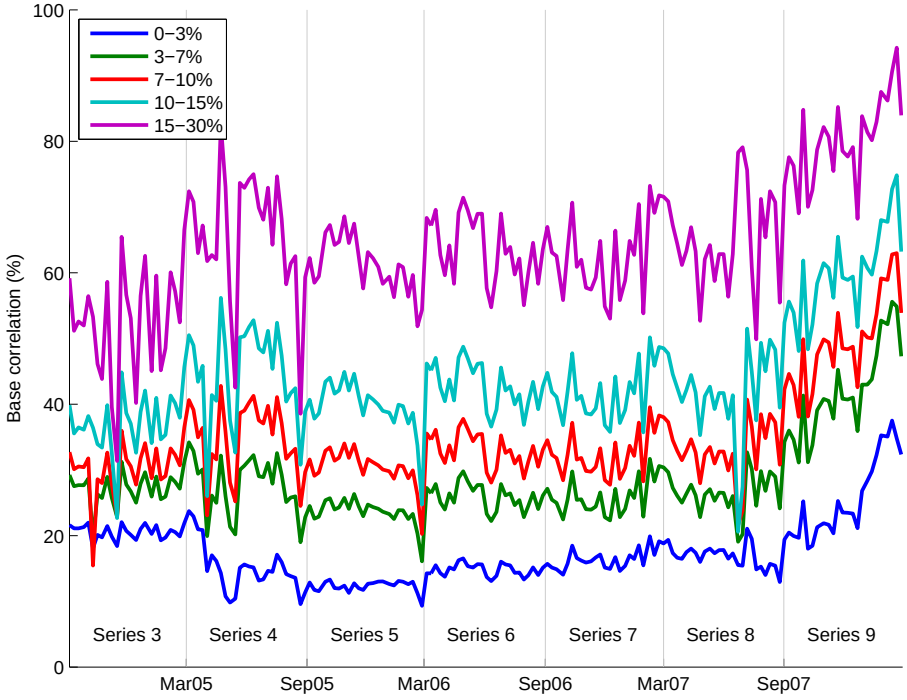
But while the OFGC is indeed commonly used, there are numerous flavors to its implementation, without any market consensus. Some of these flavors have to do with what single-credit information is used: just the index level or the CDS premia for each constituent of the index, and if the latter, whether an adjustment is made to compensate for the index basis. In all cases, we may consider either the full term structure of available prices, or just prices for the specific maturity point of concern.[2]

Further flavors of the implementation relate to the portfolio model. One dimension is whether we capture fully the discreteness of the portfolio loss distribution, that is, the fact that losses occur only in amounts corresponding to each constituent's weight in the index. A second dimension, related to what CDS data we use, is whether we capture the heterogeneity of spreads across the constituents or rather assume that all constituents have precisely the average spread. We will examine two cases: in the *granular* model, we capture the full discreteness and utilize all of the individual constituent spreads, though

[1] See Couderc (2007).

[2] All of our results will reflect the single maturity choice, as interestingly, this specification fits the data much better, at least up until the last year, than the full term structure alternative.

Figure 1: Base correlations. CDX North America Investment Grade tranches. Large pool model



we make no adjustment for the index basis; in the *large pool* model, we approximate the distribution by assuming that losses are continuous and that the constituent spreads are homogeneous.

Applying the model to actual prices, the first observation is that regardless of the flavor of implementation, the single correlation parameter is inadequate to describe more than one tranche. So we simply adopt the convention that each tranche is associated with a distinct correlation value. There are particular flavors here as well, and we adopt the conventional base correlation framework.[3] In the end, we describe the prices of the standard tranches, not by a single correlation parameter, but by a correlation curve; typically, on any day, the curve rises with increasing seniority, as can be seen in Figure 1.

The base correlation framework was first widely adopted in 2005, when its predecessor, the compound correlation framework, proved unable to describe market prices, particularly that of the 3-7% tranche on the North American index. In 2008, depending on the implementation flavor, the base correlation framework either is unable to price all tranches, in particular the senior 15-30% tranche, or can calibrate to tranche prices but only with unrealistically high values for the correlation.[4] Thus, the base correlation framework may have bought three additional years for the OFGC model, but the market

[3] See Finger (2004) for a full description of the correlation frameworks and of the standard model flavors.

[4] Indeed, some dealers actually quoted correlations over 100% during March 2008. It is not clear what these quotes signify, other than that the model is unable to describe the prices in the marketplace.

has reached a point where the model now fails the prerequisite discussed above. Nonetheless, it is still worthwhile to define some real empirical tests on the model applications, first to evaluate the different flavors of the OFGC in the past, and second to establish benchmarks for future models.

Correlation of correlation

We apply our tests to the standard tranches on the CDX North America Investment Grade index. Our data spans Series 4 through 9 of the index, and includes each series only for the time it was the most recent, or On-The-Run, series. Series 4 was first released in March 2005; from this time until September 2005, our data on the tranches and index reference this series. From September 2005 until March 2006, Series 5 was On-The-Run, and our data reference this series. The most recent series in our sample is Series 9, which was On-The-Run from September 2007 until March 2008. We note that no default has occurred on any of these series while it was On-The-Run: our data does not include any actual defaults. Series 4 was On-The-Run during the downgrades of Ford and General Motors, while Series 8 and 9 were On-The-Run during the last year's market upheaval.

Our first task is to calibrate the two models to the standard tranche prices over a long history. In the case of the large pool model, on each day, we observe the spread on the CDX index and the market prices of the standard tranches, and infer the base correlation curve. For the granular model, we observe the spreads on each index constituent, along with the tranche prices, and infer a second base correlation curve. Thus for each of the two models, we construct five time series for base correlation, corresponding to each of the five standard tranches. We plot the five series from the large pool model in Figure 1. Notably, the base correlation for the senior (15-30%) tranche approaches 100% in the spring of 2008; the granular model (not shown) is not able to calibrate to the senior tranche during this time.[5]

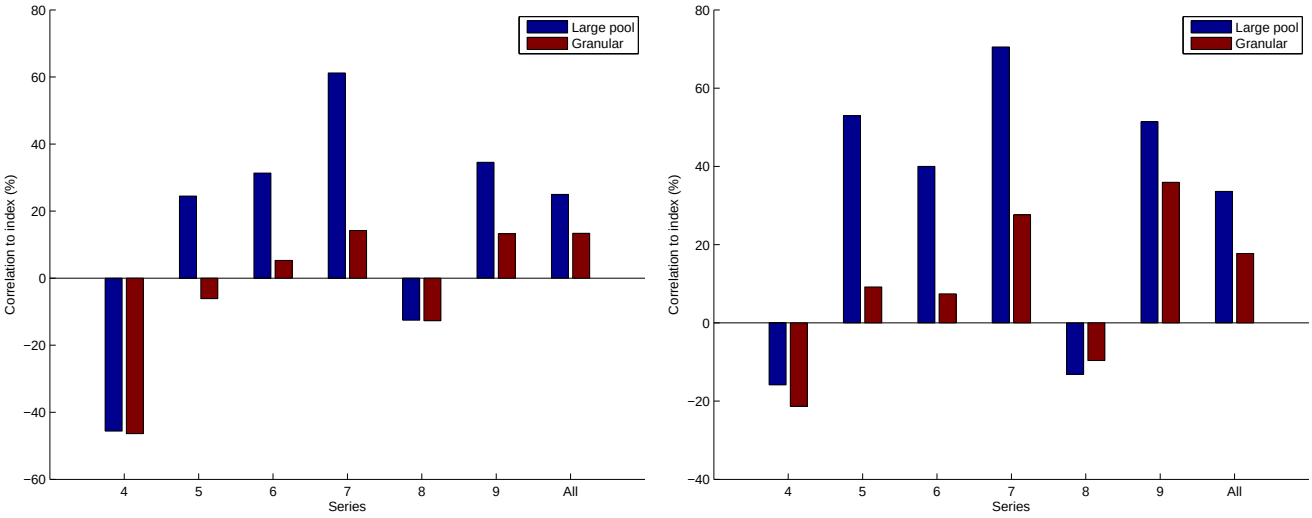
Recalling our model applications—hedging and risk aggregation—it is desirable that movements in the base correlations through time are uncorrelated with movements in the underlying credit spreads. For hedging, a strong correlation implies there is a sensitivity of the tranches to credit spreads that we are not capturing with our model. In other words, we are leaving some amount of credit sensitivity unhedged. Likewise for risk aggregation, the risk to movements in the base correlation will be a risk that we consider as idiosyncratic, a statement rendered invalid if the base correlation is driven partly by spreads.

Our first test, then, is to calculate the correlation between changes in base correlation and changes in the index spread, and to compare this across the two model formulations. We present results in Figure 2 for two of the standard tranches. In the figure, we examine weekly changes in the base correlation and spread, and calculate the correlation for each series, as well as over the entire sample period. For both tranches, the granular model produces more desirable results, in that its base correlations are close to uncorrelated with the index.

As a means to extract a good risk factor, that is, one

[5] Additionally, the granular model was not able to calibrate to the full set of market tranches during the Series 6 period either, failing here with the 7-10% tranche.

Figure 2: Correlation of base correlation changes with CDX changes. CDX 0-3% tranche (left) and 3-7% tranche (right)



idiosyncratic to other sources of risk, the granular model does appear superior, possibly simply because it employs more information. Another lesson is that the statistical properties of the two versions of base correlation are different, and so for risk analysis, it is important to apply consistently base correlations extracted from the appropriate model.

Finally, the presence of non-trivial correlations for both models indicates that there is more going on in the real world than our static OFGC model can capture. Some link between the underlying names and the tranches is not being captured, and any truly dynamic model should seek to address this.

Testing hedges

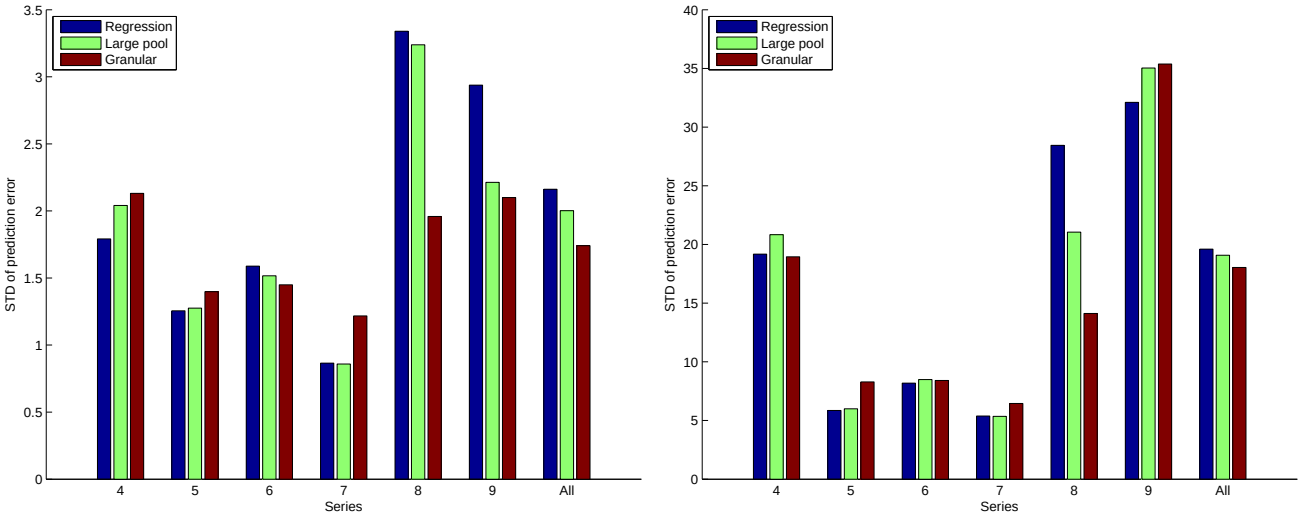
Testing the statistical properties of the base correlations, while illuminating, still does not get to the heart of the hedging issue. To test the hedges, we set up the following experiment:

1. On a given day, we calibrate our model to the standard tranche prices.
2. Over the next five days, we assume perfect foresight on the underlying credits. In the case of the large pool model, we assume we know the future index level; and in the case of the granular model, we assume we know the spreads for each underlying name.
3. Assuming that base correlations stay fixed, we apply the model with the future spreads to arrive at a predicted tranche price.
4. We compare this prediction to the actual future tranche price.

We repeat this exercise over our entire sample.[6]

[6]Of course, in reality, we would not have perfect foresight on the underlying names, and would thus construct hedges based on an assumed credit move, relying on the hedge ratio being constant for different moves. Morgan and

Figure 3: Hedge test prediction errors. CDX 0-3% tranche (left) and 3-7% tranche (right)



For comparison, we also examine a simple regression model: on each day, we regress tranche price changes on the prior six months of credit index changes; we then use this regression, along with perfect foresight on the credit index, to predict the future tranche price.

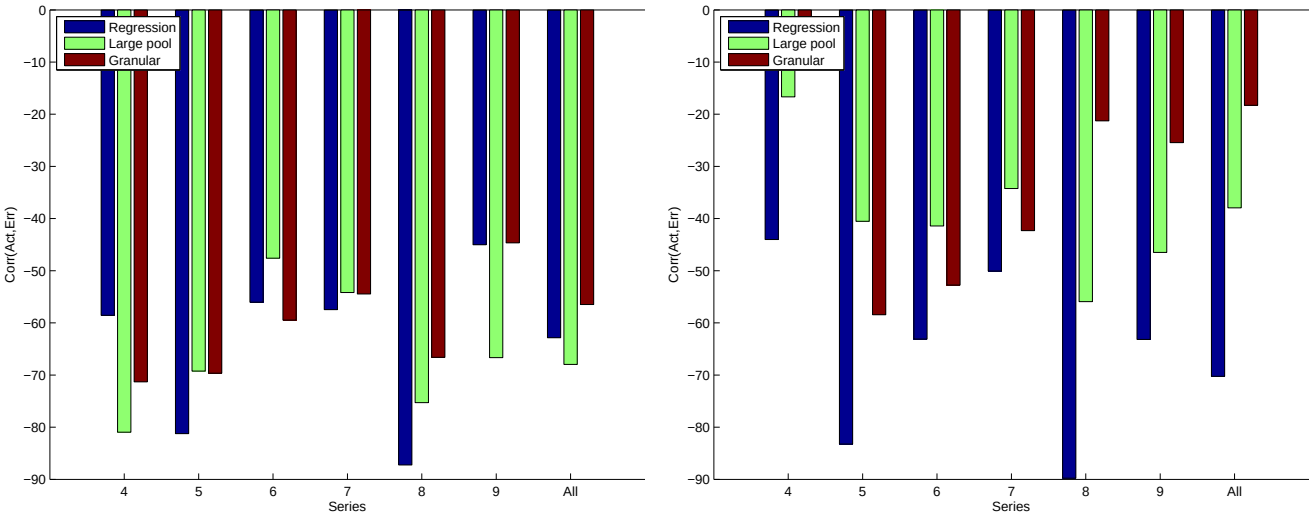
In Figure 3, we display the standard deviation of hedging errors, as measured by the difference between predicted and actual tranche prices. Most striking is that the simple regression, though it does produce the largest errors, does not fare that badly. Second, the overall level of the errors is quite large: the errors represented in the figure are generally half as large as the overall standard deviation of tranche changes, so much of the overall tranche moves is left unpredicted.

The granular model does outperform the others overall, but not for every series. It appears to provide the greatest benefit in Series 8, where this year's crisis levels of volatility began. In benign periods (Series 5 through 7), it shows little advantage. For the more senior tranches, the granular model fares even worse; the calibration difficulties mentioned before produce very large prediction errors, while the large pool and regression models are more robust.

Going beyond the error magnitudes, another general observation is that for all of the models, the predicted tranche price changes are directionally correct but tend to be too small. The implication of this is that the models systematically underhedge the actual tranche moves. We can measure this tendency by examining the correlation of the prediction errors to the actual tranche changes. We report this statistic in Figure 4. These correlations are strongly negative, indicating that the tranches in general show stronger sensitivity to credit moves than any of our models are able to capture. There is relatively little difference

O'Kane (2005) test this approach, though only on Series 4. We chose not to test this way, as it would mix the error from this linearization with the error from the pricing model itself. In a sense, ours is a pure test of the presumed relationship between underlyings and tranches, rather than of a specific hedging strategy.

Figure 4: Correlation of prediction errors with actual tranche moves. CDX 0-3% tranche (left) and 3-7% tranche (right)



between the models for the 0-3% tranche, but the underhedging problem is particularly pronounced for the regression model on the 3-7% tranche. It took a while, but we finally see a weakness in what should have been an easy model to reject.

If we factor in costs—numerical complexity and the amount of information required—it is hard to argue adamantly for the granular model. We would expect a greater performance benefit for tracking so much information. One can argue that it takes a granular model to even produce tranche sensitivities to individual names, but the hedging results here should call such sensitivities into question.

Is it really B-S?

At one point, the OFGC was referred to as "the Black-Scholes of portfolio credit". For those who helped develop the model, this was certainly a flattering moniker, but it was premature at best. What the OFGC does have in common with Black-Scholes is that there is a natural way to abuse the model (a volatility surface or a correlation curve) in order to match market prices unexplained by the fundamental model assumptions. One could also argue that as a standard for communicating prices and illuminating the market, the models played a similar role. But the great accomplishment of the Black-Scholes framework—a dynamic strategy by which one can, subject to a set of assumptions, replicate a derivative payoff—is nowhere to be seen in the OFGC model. There is nothing fundamental about the OFGC that says that its hedges should work, and so we must test empirically.

Unfortunately, in empirical tests, the OFGC shows little performance benefit over a simple regression model. Moreover, there appears to be a dynamic link between tranche prices and credit spreads that the model does not capture. We have seen that throwing more information into the OFGC (that is, moving from the index to the individual spreads) pro-

duces only marginal benefits. It is doubtful that other tweaks to the static framework will yield much; we need a model that recognizes that the market moves.

The general lesson here is that a model being mathematically aesthetic and calibrating—albeit through some contortions—to all market prices does not make the model good. The right tests are on the way the model is used: to hedge and to explain aggregate risks. It is these tests we should be applying, whether to throw the old models out or to ring in the new ones.

Further reading

- Couderc (2007). Measuring Risk on Credit Indices: On the Use of the Basis. *RiskMetrics Journal*, **7**(1): 61–88.
- Finger (2004). Issues in the Pricing of Synthetic CDOs. *RiskMetrics Journal*, **5**(1): 21–40.
- Finger, C. (2005). Eating Our Own Cooking, RiskMetrics Research Monthly, June.
- Morgan, S. and O'Kane, D. (2005). Base Correlation Deltas: An Empirical Investigation. Lehman Brothers Quantitative Credit Research. November.