Methodology for Testing the Level of the EDF[TM] Credit Measure[1]

Matthew Kurbat and Irina Korablev[2]
Moody's | KMV[3]
San Francisco
Technical Report #020729
Revised: 08/08/02

Abstract

Users of default risk models naturally wish to verify that the average level of predicted defaults tracks realized defaults over time. A straightforward approach to level validation is to compare mean predicted default rates with empirically observed default rates to see that they match, within limits of sampling error. However, the skewed distribution of mean default rates implies that the straightforward approach is inappropriate. Ignoring this fact may cause predicted default rates to *erroneously* seem too high, an error that is greatest for large, low-risk firms. This paper describes how to perform level validation in the face of such skewness. We then compare the levels of the EDF credit measure with realized default rates from 1991-2001, and find that the predicted and realized levels match within limits of sampling error. Our primary focus is on US large firms where the presence or absence of default can itself be more readily validated.

1. Introduction

The output of Moody's | KMV Credit Monitor and Credit Edge is the EDF credit measure. An EDF value is a quantitative measure of credit quality. More specifically, an EDF value estimates the probability of default for a given firm. This probability estimate is bounded between 0.02% (for an EDF value of 0.02) and 20% (for an EDF value of 20).[4] For an overview of the EDF credit measure, see Crosbie and Bohn (2001).

The New Basle Capital Accord states that banks must have a robust system in place to validate the accuracy of estimated default probabilities.[5] An important part of any such validation is to verify that the average level of predicted defaults tracks realized defaults over time. Note that this issue, which we will call *level validation*, tells us how well a model is *calibrated* (whether or not a model's *average* predicted level of defaults is correct). The issue of calibration is distinct from the *power* issue of whether or not a model can discriminate *individual* good firms from bad firms (e.g., Stein, 2002). A model can have great power but be poorly calibrated, or vice-versa (although it is generally easier to improve calibration than power). This paper focuses on level validation and calibration, although other work has shown that the EDF credit measure is also a powerful predictor of default (e.g., Kealhofer, 2002a).

A straightforward approach to level validation is to compare mean predicted default rates with empirically observed default rates to see that they match, within limits of sampling error. However, because mean default rates have a distribution that is not normal but is in fact skewed, this straightforward approach induces systematic biases that hamper attempts to conduct level validation. This paper outlines an approach that allows us to perform level validation in the face of such skewness. We then use the approach to compare the levels of the EDF credit measure with realized default rates from 1991-2001, and find that the predicted and realized levels match within limits of sampling error. Our primary focus is on US large firms where the presence or absence of default can itself be well validated, but we also discuss other cases.

The remainder of this paper is laid out as follows:
- Section 2 lays out the statistical groundwork, using simulated data to focus on the effects of sample size and most importantly skewness in default distributions. The main point of section is that ignoring skewness in default distributions will tend to *erroneously* make predicted default rates *seem* too high, and that this spurious bias will be greatest for large, low risk (e.g., investment grade) firms.
- Section 3 reviews some methodological details needed to apply default prediction methods from section 2 to actual data.
- Sections 4 and 5 apply our approach to real data, in situations where the default data are most solid. We find here that EDF-predicted and actual default rates track each other well over time.
- Section 6 reviews some important, related issues. One is that skewness is minimized by using a very long history of defaults (one or more decades worth of data). A second issue is that differences in power between two well-calibrated models can, surprisingly, lead to spurious apparent differences in calibration.

2. Statistical Issues in Validation

This section lays the statistical groundwork, using simulated data to focus on the effects of sample size and most importantly skewness in default distributions. The **most important point** of this section is that ignoring skewness in default distributions will tend to *erroneously* make predicted default rates *seem* too high, and that this spurious bias will be greatest for large, low risk (e.g., investment grade) firms.

On the surface, level validation consists of the following straightforward steps:
- Begin with a set of N firms at the start of a time period.
- Each firm is estimated to have default probability p_i over that time period.
- The mean predicted number of defaults (P) is simply the sum of those N default probabilities ($P = \Sigma \, p_i$).
- At the end of the period, we count the number of actual defaults A over the period in the set of N firms.
- If the predicted P and actual A default counts are close, then the model is validated, otherwise it is not.

In practice, validation is not quite so straightforward for at least two reasons: (1) an actual default rate (sample) estimates the "true" underlying default rate (population) with measurement error due to sample size constraints, making it sometimes difficult to know how "close" is close enough; (2) as we will demonstrate, the distribution of the sample mean is not normal, but skewed. After reviewing notation, we then discuss each of these issues in turn.

2.1 NOTATION

We use the following notation in this section:

P = probability of default
EDF = 100*P (EDF values are scaled from 0-100, P is scaled from 0-1)
D = Default count over a time period (i.e., number of defaults)
N = Number of firms in a sample over a time period
R = Default rate over a time period (i.e., D/N).

2.2 SAMPLE SIZE CONSTRAINTS AND MEASUREMENT ERROR

Whenever we look at the results of a default study, we see reports of statistics. A statistic is a random variable that is intended to measure a true underlying value. Because we only observe the statistic and not the true value, it is natural to ask how close one is to the other. In this section we examine this measurement error assuming a normal distribution as an approximation (we later relax this assumption).

The relative rarity of default events can pose a challenge for validation: to be confident that the statistic is close to the true value, we need a large amount of data. For example, given 1000 uncorrelated firms with true population default probabilities of 0.001, we expect to see (0.001)*(1000)=1 default over a one year time horizon. However, in a given year a single firm either defaults or it does not. So, although we expect a default rate of 0.1%, for a given firm we observe a default rate of 0% (close to the expected rate) or 100% (far from the expected rate). When the default probability is small, the standard deviation around the expected number of defaults D tends to be large relative to D. For N uncorrelated defaults with the same population default probability P, this standard deviation is S=sqrt(N*P*(1-P)).[6][7] Continuing with our hypothetical 0.1% expected default rate and 1000 firms, S = sqrt(1000*(0.001)*(0.999)) ≈ 1.00. Assuming a normal distribution as an approximation, we can then obtain an approximate 95% confidence interval of D± 1.96S = 1 ± 1.96*1.00 ≈ [-1, 3], or [0, 3] if we truncate the lower end at zero.

The next table illustrates this thinking for various default probabilities (P) and sample sizes (N):

P	N	Defaults	St.Dev.	Lower	Upper	CV
0.001	100	0.1	0.32	0.0	0.7	320.00
0.001	1000	1.0	1.00	0.0	3.0	100.00
0.001	10000	10.0	3.16	3.8	16.2	31.60
0.01	100	1.0	0.99	0.0	2.9	99.00
0.01	1000	10.0	3.15	3.8	16.2	31.50
0.01	10000	100.0	9.95	80.5	119.5	9.95
0.1	100	10.0	3.00	4.1	15.9	30.00
0.1	1000	100.0	9.49	81.4	118.6	9.49
0.1	10000	1000.0	30.00	941.2	1058.8	3.00

In this table, "Defaults" is the mean predicted number of defaults, "St.Dev." is the standard deviation, and "Lower" and "Upper" bound the 95% confidence interval for the predicted number

---

[6] Because of correlation between defaults, and non-normality in smaller samples, basing a confidence interval on this idealized approximation is subject to error (more on this below). Also, EDF values must inevitably be estimated with some degree of error, which makes prediction intervals wider than in-sample confidence intervals by some amount. For an elaboration of the prediction interval versus confidence interval distinction, see p.399 of DeFusco, *et al.* (2001).
[7] Note that predicted default *counts* have mean NP and standard deviation sqrt(NP(1-P)), while predicted default *rates* (counts divided by sample size) have mean P and standard deviation sqrt((P(1-P))/N).

of defaults. "CV" is the coefficient of variation (100*{st.dev/mean}), which in this context is essentially a measure of the degree to which the mean is likely to be measured with error – the larger the CV, the more error-prone the default count estimate.

Two points are evident from looking at these results, especially the CV values:[8]
- For a given level of default risk, we can measure that risk more accurately with a larger sample size (e.g., in the cases with P=0.1, larger sample sizes N correspond to lower CV).
- For a given sample size, we can measure the default rate more accurately for firms with higher default risk (e.g., in the cases with sample size N = 10000, higher P corresponds to lower CV).

To be concise: the more data the better, especially for low risk firms.

If defaults were independent, then the values in the above table would provide us with a rough sense of what a reasonable range of observed default count is for a given predicted default count. For example, given a CV of 30, we can reasonably expect the realized default count to be roughly within plus or minus 60% of the predicted default count 95% of the time (as we can see in the table, a slightly wider interval is appropriate for lower default probabilities). Unfortunately, the assumption of independence is problematic here, as we shall see shortly.

2.3 SKEWED MEAN DEFAULT RATE DISTRIBUTIONS
The results just shown with a normal approximation to find confidence intervals around mean default rate estimates are useful: they illustrate that increasing sample size improves such estimates. However, the distribution of mean default rate estimates is positively skewed, not normally distributed. Consequently, there is a larger probability of observing a realized default rate below the mean value than one above the mean value, but there is a greater possibility of observing a realized default rate that is *far* above the mean value than one that is far below the mean value.

2.3.1 Why are mean default rate distributions skewed?
Two reasons for skewed default rate distributions are: (1) sample size is small relative to default rate; (2) correlated defaults. We discuss each in turn.

2.3.1.1 Small sample size relative to default rate
For N independent firms with the same default probability P, default counts have a binomial distribution with parameters N and P (default rates are simply counts divided by N). A Binomial distribution with parameters N and P is well-approximated by a normal with mean NP and variance NP(1-P), but only if NP(1-P) > T, where T is typically a value in the range [5,10].[9] Sample sizes used in year-by-year validation may not be large enough to meet this restriction. For example, given 1000 firms with EDF values of 0.10, NP(1-P) ≈ 1.00 < 5.

The figures below demonstrate this result using simulated data. The figures show the simulated default distribution[10] for the case of 1000 firms with EDF values of 0.10 (on the left) and 5000 firms (on the right). Sample sizes in this 1000 - 5000 firm range do arise in practice. Note that NP(1-P) ≈ 1.00 < 5 (so we expect skewness) on the left where skewness is pronounced, and

---

[8] See Cantor *et al.* (2001) for further discussion of these points.
[9] e.g., Johnson, Kotz, and Kemp (p.114, 1993).
[10] We used 10,000 trials for each simulation.

NP(1-P) ≈ 5.00 (so we expect only slight skewness) on the right where skewness is relatively minor.[11]
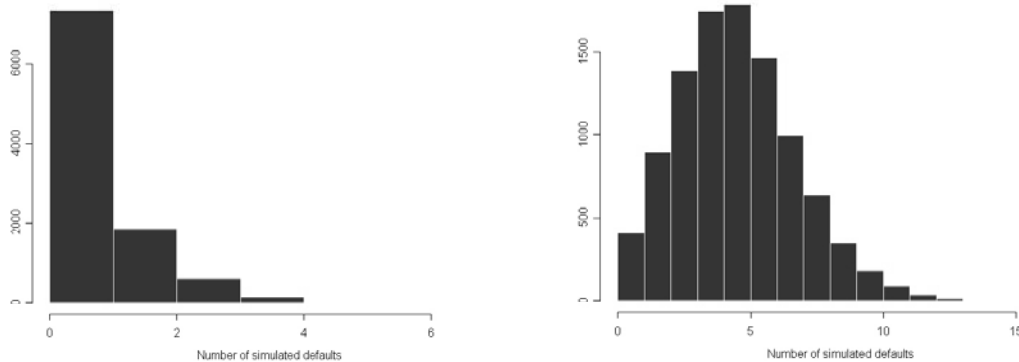
2.3.1.2 Non-independent defaults

The previous discussion assumes that default events are independent. In fact, they are not. Greater inter-dependence of defaults leads to more skewed default rate distributions.

We can view default as occurring when the value of a firm falls below a certain threshold value. On this view, the random event of default is caused by a more fundamental random event, namely the drop in firm value. Using this view, default correlations can be inferred from correlations at the level of firm values, rather than as default correlations. Why do we do this? Because defaults are relatively rare, making joint defaults even less common. In fact, the historically observed joint frequency of default between any two companies is usually zero. Grouping firms allows us to estimate average default correlation in the group using historical data, but the estimates obtained this way are very inaccurate. The Moody's | KMV approach to modeling default correlation as a consequence of asset correlation is presented in Kealhofer and Bohn (2001). We also have an asset correlation model, known as Gcorr[TM], developed specifically for modeling asset correlations (see, e.g., Zeng and Zhang, 2001).

The median asset correlations for groups of firms examined later in this paper are all in the 0.1-0.2 range, though correlations between specific pairs of firms (e.g., those in the same industry) can be considerably higher. The positive asset correlations are similar to the positive correlations observed between equity returns for pairs of firms, except that equity correlations are higher than the corresponding asset correlations due to effects of leverage on the former.

To demonstrate the impact of correlation on defaults, we simulate default count distributions below. We show results using several values for asset correlation, EDF value, and sample size to illustrate the effect of these parameters on the relative magnitudes of the mean and median of the default count distributions:[12]

---

[11] A scale-invariant measure of degree of asymmetry of a distribution is the *coefficient of skewness*. For the binomial distribution, this measure is $(1-2P)/(NP(1-P))^{\frac{1}{2}}$ The general form of this measure is μ_3/σ^3, where μ_3 and σ are the third central moment and standard deviation, respectively, of the distribution. See Evans, Hastings, and Peacock (1993) or Stuart and Ord (1994, p.108-110). To forestall confusion, note that this measure is scale-invariant, and so is *not* the more typically quoted skewness that is the 3rd moment about the mean, which for the binomial is NP(1-P)(1-2P).
[12] We used 10,000 trials for each simulation.

| | | 1000 firms | | | 10,000 firms | | |
Correlation	EDFvalue	Mean	Median	Ratio	Mean	Median	Ratio
0.0	0.1	1.0	1	**1.00**	10.0	10	**1.00**
0.0	1.0	10.0	10	**1.00**	100.1	100	**1.00**
0.0	10.0	100.0	100	**1.00**	999.6	1000	**1.00**
0.1	0.1	1.0	0	**0.00**	9.8	6	**0.61**
0.1	1.0	9.9	7	**0.71**	100.5	71	**0.71**
0.1	10.0	99.9	89	**0.89**	1002.0	883	**0.88**
0.2	0.1	1.0	0	**0.00**	10.2	3	**0.30**
0.2	1.0	10.0	5	**0.50**	99.4	46	**0.46**
0.2	10.0	99.5	74	**0.74**	1001.5	758	**0.76**
0.3	0.1	1.0	0	**0.00**	10.2	1	**0.10**
0.3	1.0	9.9	3	**0.30**	99.9	27	**0.27**
0.3	10.0	102.0	63	**0.62**	1006.2	629	**0.63**

"Mean" and "median" here are mean and median numbers of defaults for each type of simulation,[13] while "ratio" is simply median/mean. Because the median is less than the mean for a right-skewed distribution, we use the ratio (median/mean) as a measure of skewness here. Note the following stylized facts about default distributions based on the patterns in these ratios:
1) Skewness increases as asset correlation increases;[14]
2) Skewness increases as default risk decreases;
3) Skewness increases as sample size decreases.[15]

The points about default risk and sample size replicate those we saw in the no-correlation case previously. The observation about asset correlation is most important here because it shows that skewness is more pronounced for realistic (i.e., positive) levels of correlation than it is for the zero correlation assumption typically used in literature on this problem. Note that the sample sizes involved (1000, 10000) are fairly large, and are representative of those seen in some empirical studies.

2.3.1.2.1 Why do we have to find predicted default counts by simulation?

In certain special cases (such as the one shown next), it may be possible to find predicted default counts by solving one or a small number of equations. However, in the general case, we use simulation to find predicted default counts because there is no such closed form solution we know of that provides the answer.

2.3.1.2.2 Results as the sample size becomes arbitrarily large

As shown above, increasing the sample size used in validation studies decreases the skewness of default distributions. However, positive asset correlation creates limits on this effect that we do

---

[13] Again, we can convert from a default count to a default rate by simply dividing the count by sample size.
[14] We also note that effective sample size decreases as asset correlation increases. For example, looking at the case with 1000 firms and EDF values of 10, the absolute values of differences between sample means and population means for correlations of (0.0, 0.1, 0.2, 0.3) are (0.0, 0.1, 0.5, 2.0) respectively. Note that these error magnitudes are increasing as correlation increases, as we would expect because effective sample size is decreasing.
[15] For given EDF value and asset correlation, higher sample sizes tended to correspond to equal or higher ratios, but not always. Why are there counter-examples to this trend? For several reasons: (1) because medians are reported in integer values, we expect a certain degree of rounding error in the median values, particularly for the 1000 firm cases; (2) given the relative rarity of defaults in many of these simulations, we expect some degree of rounding error even using samples of these sizes.

not see in the zero correlation case. To see this, we make use of some results from Vasicek (1991). Vasicek examined the loss distribution of a loan portfolio with the same default probability P for each loan, and constant pair wise asset correlation ρ. Looking at the case where the number of loans in the portfolio goes to infinity, he finds the mean and median of the resulting default rate distribution[16] to be

$$\theta_{mean} \quad = P$$
$$\theta_{median} \quad = N(N^{-1}(P)/(1-\rho)^{\frac{1}{2}})$$

We can use this result to find the ratio of median over mean ($\theta_{median}/\theta_{mean}$) for a portfolio with an infinite number of firms. Calling this the "asymptotic" result, we compare it with the corresponding ratios from the previous table for sample sizes of 1000 and 10000, as shown below:

		Median/mean by number of firms		
Correlation	EDFvalue	1000	10000	Asymptotic
0.0	0.1	1.00	1.00	1.00
0.0	1.0	1.00	1.00	1.00
0.0	10.0	1.00	1.00	1.00
0.1	0.1	0.00	0.61	0.56
0.1	1.0	0.71	0.71	0.71
0.1	10.0	0.89	0.88	0.88
0.2	0.1	0.00	0.30	0.28
0.2	1.0	0.50	0.46	0.46
0.2	10.0	0.74	0.76	0.76

The results in this table can be described by the following stylized facts:
- When asset correlation is positive, the ratio of median/mean remains less than one even for our asymptotic case. *This implies that no sample size is large enough to remove skewness from default rate distributions that have positive asset correlation.*
- For EDF values at or above 1, we observe no significant reduction in skewness by using more than 1000 firms, because ratios for the 1000 firm and asymptotic case are essentially the same.[17]
- For EDF values below 1, we do observe significant reduction in skewness by increasing sample size above 1000 firms, although this benefit of increased sample size stops by 10000 firms.[18]

2.3.2 Default distributions are skewed because debt portfolio loss distributions are skewed

It sometimes helps to understand why mean default rate distributions are skewed by thinking of a default distribution as a special case of a debt portfolio loss distribution. In this case, the portfolio holds N equal-sized positions, and a loss variable takes a value of 0 (no default) or 1 (default) for each of the N positions. Here again the default count is the sum S of the N loss variables, and the default rate is S/N. Framing the problem in this way can help because we know loss distributions are highly skewed for debt portfolios (e.g., Kealhofer and Bohn, 2001), so default distributions must also be highly skewed because they are a special case.[19]

---

[16] Here, "N" is the cumulative normal distribution function, and "N^{-1}" is the inverse cumulative normal.
[17] The fact that the ratios are sometimes slightly larger for the 1000 firm case than for the asymptotic case is due to sampling noise.
[18] The same sampling noise point that applies to the 1000 firm case also applies to the 10000 firm case.
[19] See also Kealhofer, Kwok, and Weng (1998).

2.3.3 Why should we care that default rate distributions are skewed?

Why does any of this matter? It matters because, if a mean predicted default rate is used in an attempt to validate an accurate default risk model, the model's predictions may erroneously appear to be biased because the validation procedure is incorrect.[20]

Consider the following hypothetical example, which we chose because it is representative of cases we see in actual validation. Say we take a case where we have 1000 firms, each with EDF values of 1.00 and pair wise asset correlations of 0.15, then simulate 10,000 replications of this case and examine the resulting default distribution. What we get, as shown in the figure below, is a highly skewed distribution:[21]

Number of simulated defaults
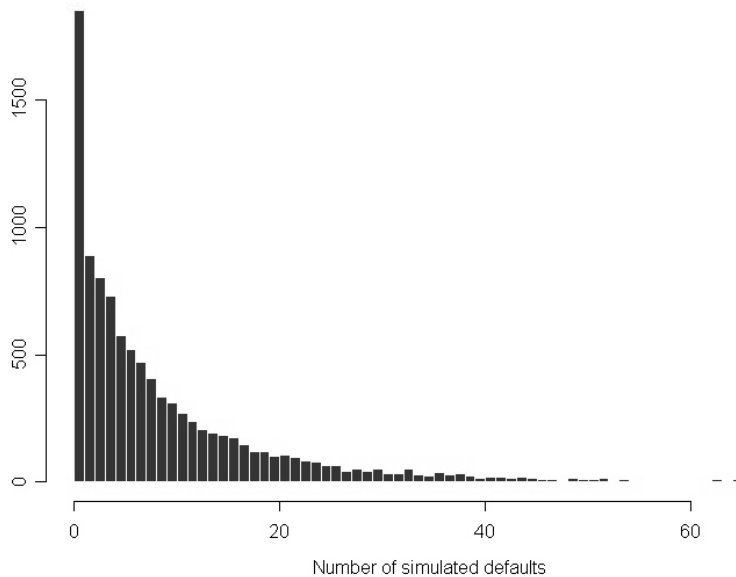
This distribution has a median (50th percentile) of 6 defaults and a mean of 10 defaults (1000*0.01).[22] In this case the mean is at the 67th percentile, so if we evaluate the model based on comparing observed results with the predicted mean, the model would erroneously appear to be too high in 66 out of 100 cases, or two out of every three. *In other words, the average level of default predictions would appear to be too high, not because the model was incorrect, but because the methodology used to check the model was incorrect.*

Each observed default rate is a single point drawn from a distribution. If we could draw a large number of points from the distribution we would arrive at a sense of what results are typical. When we have only one result per group, we don't know whether our sample mean estimate happens to be one of the lower values or one of the higher values. So, this spurious problem

---

[20] To underscore a point, note that by "bias" here we do *not* mean "expected value," as that is what the mean default rate literally represents.

[21] To better fit on the page, this picture excludes the 72 out of 10,000 trials that had more than 70 defaults, the highest of which had 149 defaults (this "long right tail" of the distribution was not visible in the graph because these 72 cases were spread out over the range 71-149).

[22] The 25th and 75th percentiles of this distribution are 2 and 13 defaults, respectively.

might not be noticeable for any single data set, but if we were to compare model predictions with actual defaults separately over a number of different years, the pattern would likely become more salient.

2.3.3.1 *Why we want to use the median of the simulated distribution as the predicted default rate or count*

What we expect to see when comparing *accurate* predicted default counts with actual counts is for predicted to lie above actual roughly 50% of the time, and below actual roughly 50% of the time, due to sampling variability. Use of mean predicted defaults does not do this for us, as just shown. However, taking the median of the simulated distribution provides us with a predicted value that *by definition* has 50% of the observations each above and below. So, assuming the model is well calibrated, taking the median of the simulated distribution will give us what we want.

2.3.3.2 *Why can't we just use median EDF values for predicted default rates instead of taking the median of the simulated distribution?*

It is incorrect to simply use median EDF values for predicted default rates. For example, consider the preceding simulation. In that case, we had 1000 firms, each with EDF values of 1% and pair wise asset correlations of 0.15. Here our median and mean EDF value are both the same (1%), but this number differs from the simulated median default rate of 0.6%. In sum, the median EDF value is not the same as the median of the simulated distribution.

2.3.4 What do these results imply for practical validation?

In the previous section, we found that skewness increases for:
- Firms with higher average asset correlation
- Firms with lower EDF values
- Smaller samples

Importantly, this result implies that, if we incorrectly ignore skewness and focus on mean default rates, then EDF values will *appear* to over-predict default rates the most for larger (due to higher asset correlations), investment-grade firms. Also, this spurious over-prediction will be even more pronounced if the sample of firms and defaults we use is not as large as possible. *These points are critical for any organization that attempts to perform its own validation.*

2.4 SUMMARY AND IMPLICATIONS

The key points of this section are:
- The mean is only a good measure of the center of a predicted distribution that is symmetrical. Default distributions are notably skewed, making the median of the simulated default distribution a better measure of the center of that distribution to use for validation. Using the mean instead will make estimates appear to be too high.
- Skewness, and the resulting gap it causes between mean and median, is larger for firms with higher asset correlations and lower EDF values, and for smaller sample sizes.

Some implications of these points are:
- Validations based on overly small samples will be less useful, and in fact may do more harm than good as default rates may appear to be incorrect when in fact results are too noisy to be meaningful.
- Because skewness is most pronounced for firms with higher asset correlations and lower EDF values, the apparent bias that would result from erroneously using the mean predicted default rate will be largest for big, stable firms.

3. Empirical Methodology

The previous section reviewed some general points about level validation, illustrated by *simulated* data. Now we prepare to test our level validation approach on *real* data. The details of how we compared predicted and actual numbers of defaults are as follows:

- **Time period.** Results are broken out by year for the period 1991-2001.[23] Defaults used to calibrate the model were from 1973 to 1995, so the period 1996-2001 is an out of sample test.[24]
- **Population.** US public non-financial[25] firms with sales above $300 million in 2001 US dollars. This threshold allows others to replicate these results because the presence or absence of defaults is easier to validate above it.[26]
- **Actual defaults**: We count the number of defaults that occurred in each year, using our proprietary default database as the source of the defaults.
- **Predicted defaults**: We find all firms meeting the above restrictions at the start of each year. The predicted mean number of defaults is simply the sum of their EDF values divided by 100.[27] Predicted median numbers of defaults are obtained via Monte Carlo simulation, using these firms' EDF values and using an average level of asset correlation in the simulation.[28] [29] Use of a single average correlation is simpler for someone who wishes to replicate this methodology. Details of the simulation of correlated asset values (and therefore defaults) are discussed in an appendix.

4. US Results

We first present results broken down by coarser levels of the EDF credit measure, then repeat the analysis for narrower ranges of the measure.

4.1 OVERALL RESULTS

We separate results into two groups: (1) EDF values less than 20; (2) EDF values equal to 20. This separation is needed because the truncation at 20 is a lower bound: such firms should have a default probability of 20 or more, so we separate results for these firms from those for non-truncated cases.

---

[23] Calendar years provide convenient units of time into which to break down the results. However, we might expect sampling variability in our results to be smallest if we choose to begin each year in say April, because annual financial statements that contain updated liability information are often first available then. It would be interesting to consider the impact of statement staleness on sampling variability in future work.

[24] Results post-1995 are in a sense more interesting because earlier years are in sample. For this reason we only include the most recent part of the in-sample period that allows us to observe results over an entire credit cycle.

[25] We use non-financial firms to be consistent with earlier power curve work (e.g., Kealhofer, 2002a), although financial firms comprise only about 10% of the total population, and less than 10% of the defaults, and so have a negligible impact on the results.

[26] Hidden defaults are more prevalent for smaller firms in part because they are less newsworthy.

[27] We use simulation to obtain the median predicted number of defaults, so why don't we use it to obtain the mean? In fact, the two converge for a sufficiently large number of replications. We compared means obtained from simulation with those obtained from sum(EDF)/100 in the simulations here, and the differences between the two were negligible.

[28] Our approach to simulation, reviewed in an appendix, is Moody's | KMV's standard approach of modeling joint default risk through asset value correlation, as discussed by Kealhofer and Bohn (2001).

[29] Using overall average correlation is not quite as good as Moody's | KMV Gcorr estimates, but still forecasts correlation relatively well (Zeng and Zhang, 2001).

4.1.1 Results for firms with EDF values below 20

The graph below displays median predicted (by simulation) and actual defaults for EDF values below 20. The median asset correlation for pairs of firms in each year was 0.167, so that value was used to simulate defaults for this sample, which contained a total of 19,278 firm-years.[30]
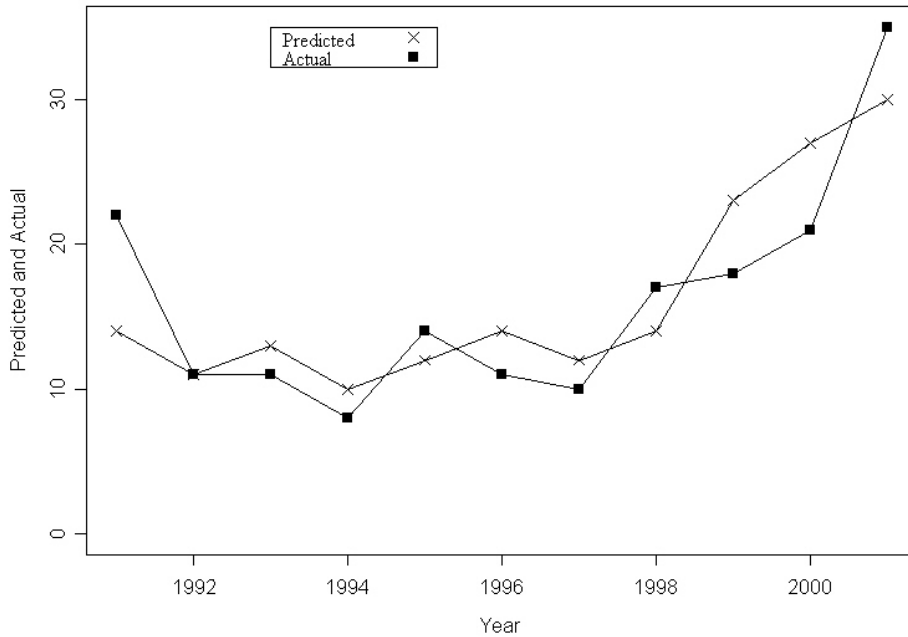

The next table presents the numbers that underlie the previous graph: number of firms, number of defaults, median and mean predicted number of defaults per year:[31]

---

[30] This count sums the number of firms in the simulation for each year across the years 1991-2001.
[31] For those that wish to look at default proportions instead of counts, it is straightforward to obtain proportions by dividing each default count by the total number of firms to obtain the default proportion. We focus here on default counts instead of default proportions in order to highlight the sampling error issues involved.

Year	Firms	Defaults[32]	Median[33]	Mean
1991	1457	22	14	21.72
1992	1482	11	11	15.96
1993	1574	11	13	17.68
1994	1667	8	10	15.12
1995	1706	14	12	18.54
1996	1816	11	14	20.57
1997	1953	10	12	18.13
1998	2028	17	14	20.18
1999	2027	18	23	32.39
2000	1812	21	27	38.06
2001	1756	35	30	40.73

The difference between median and mean predicted defaults shown in the table is worth emphasizing. To underscore this difference, we next replicate the previous graph with mean instead of median predicted number of defaults:


---
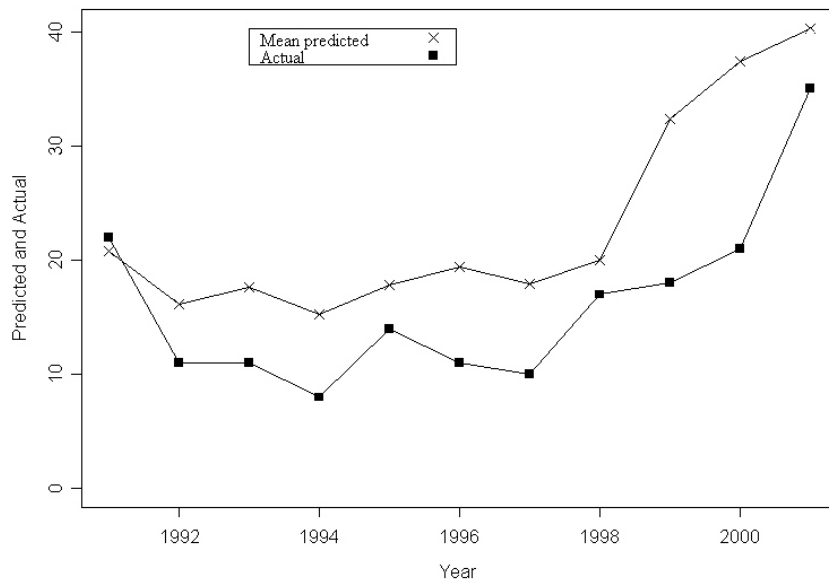
[32] Importantly, the default counts for firms of this size are larger if we also include firms with EDF values of 20, which we show below. For example, in 2001 we observed 35 firms with EDF values below 20 defaulting, and an additional 52 defaults if we add firms with EDF values of 20 at the start of the year, for a total of 87. In contrast, Moody's Investor's Services (MIS) lists 142 rated US defaults for 2001 (Hamilton, Cantor and Ou (2002)). Why the discrepancy? Most of the difference is due to the fact that the MIS list includes private firm defaults, while our list here includes only public firms. The remainder of the difference is caused by: (1) firms in the MIS list being too small to make our $300 million in sales cutoff; (2) a few financial defaults (which we excluded); (3) defaults that were not the first instance of default for an issuer in this period of distress (we included only the first instance).

[33] Simulation results for the empirical work in this paper are based on 1000 iterations; a larger number of iterations would be needed if we were estimating quantities in the tails of the default distributions.

We have the following comments on these two graphs:

- Median predicted and actual numbers of defaults track each other well, given sampling error.[34]
- Mean predicted and actual numbers of defaults do NOT track each other well. Attempts to use the mean predicted number of defaults (instead of the median of simulated defaults) will *erroneously* make the EDF measure appear to be biased upward.
- When comparing median predicted and actual default rates, the two years where actual rates exceed predicted rates the most are 1991 and 2001. Not coincidentally, these are recession years. Underestimates in such years may be the result of macroeconomic shocks that would have been unforecastable (e.g., 9/11/01) at the start of the year. In model calibration, the average level of predicted default must "split the difference" between such periods and other times when no such shock occurred, making the level slightly too low in one case and slightly too high in the other. One important point to make about unforecastable shocks (like 9/11/01) is that it is impossible to improve the model to account for them, short of having a crystal ball for a model. A second important point is that such shocks reduce the effective sample size and increase the size of confidence intervals, which to some degree renders invalid any statistical tests or confidence intervals that do not account for this issue.[35][36]

Finally, in our earlier discussion we claimed that the distribution of residuals (median predicted minus actual) should be skewed. In other words, when actual defaults are lower than predicted they will tend to be only a little lower, but when actual defaults are higher than predicted they may be more than a little higher. To assess this claim, we present the histogram of these "median predicted minus actual" results below. Although the sample size is of course small, the plot is in fact skewed to the right as we expect:

Median predicted minus actual defaults

---
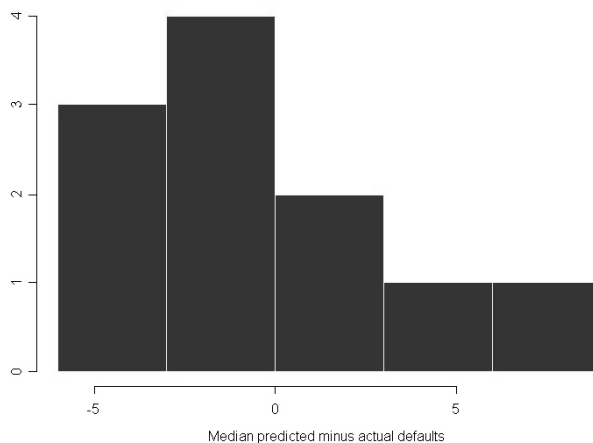
[34] As a sensitivity check (see Appendix), we repeat the median-predicted-versus-actual exercise with asset correlation values of 0.1 and 0.2 (these values form an interval around the average value of 0.167 used here). This check is useful because the level validation test is a joint test of the average level EDF values and asset correlations.

[35] We note that our simulations of median default rate used here do not include time-varying correlations to account such shocks, for the simple reason that such time-variability is too difficult to forecast over the long horizon needed. For related discussions see Cantor *et al.* (2001) and Kealhofer (2002b).

[36] Because of these macroeconomic shocks, we find it inappropriate to use our simulated distributions to attempt to place confidence bands on predicted defaults.

4.1.2 Results for firms with EDF credit measure equal to 20

The graph below displays median predicted and actual defaults for EDF values of 20. The median asset correlation for pairs of firms in each year was 0.136, so that value was used to simulate defaults for this sample of 511 firm-years.[37]
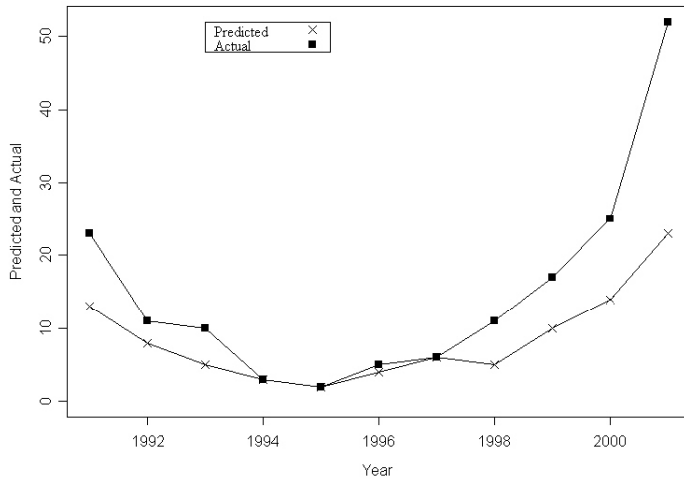
The next table presents the numbers that underlie the previous graph: number of firms, number of defaults, median and mean predicted number of defaults per year:

Year	Firms	Defaults	Median	Mean	Defaults/Median
1991	66	23	13	13.65	1.77
1992	45	11	8	9.20	1.38
1993	28	10	5	5.49	2.00
1994	15	3	3	2.97	1.00
1995	13	2	2	2.44	1.00
1996	23	5	4	4.55	1.25
1997	32	6	6	6.48	1.00
1998	26	11	5	5.13	2.20
1999	58	17	10	11.60	1.70
2000	77	25	14	15.39	1.79
2001	128	52	23	25.60	2.26

We have the following comments on this graph:
- Predicted values provide a "floor" under actual numbers, which we expect given that the EDF measure is truncated at 20. This is shown by the ratio "Defaults/Median", which takes the number of defaults for each row and divides it by the median predicted number of defaults. The fact that this ratio is always greater than or equal to 1.0 in all 11 cases suggests that the EDF values are properly calibrated, and serve as a lower bound on default frequency given their truncation at 20.

---

[37] This count sums the number of firms in the simulation for each year across the years 1991-2001.

- We again see the largest gap between predicted and actual rates for 1991 and 2001. This is partly due to the fact that the firms involved are more distressed on average in 1991 and 2001 than in other years (this is not observable here because of the EDF truncation) and possibly in part (as in the EDF value < 20 case) due to macroeconomic shocks.
- Note that mean and median predicted values here are now much closer.[38]

4.1.3 Distribution of results for a single year as an example

We present below the default count distributions for 1000 simulated iterations in each group for 2001. The left and right panels show results for EDF values <20 ("lower risk") and =20 ("higher risk"), respectively; the x-axis shows number of defaults, and the y-axis shows number of iterations for which we observed that number of defaults:
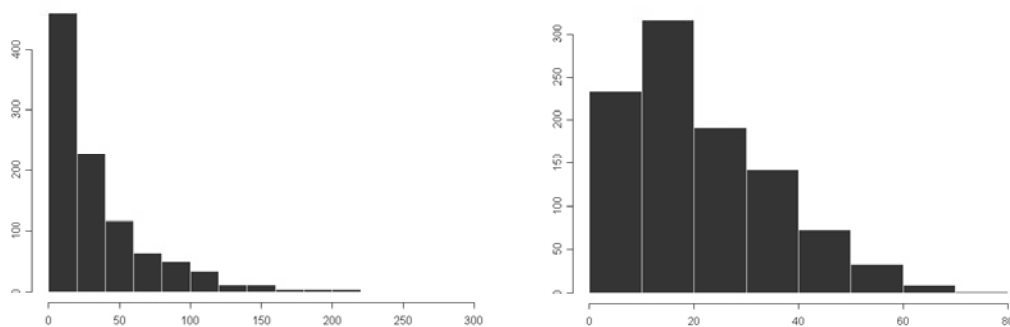
For the lower risk case on the left, the median and mean predicted numbers of defaults are 30 and 40.7, while for the higher risk case on the right, the median and mean predicted numbers of defaults are 23 and 25.5. So, the ratio (median/mean) is 30/40.7 = 0.737 for the lower risk case, and 23/25.5 = 0.902 for the higher risk case. These results highlight the fact that use of mean predicted defaults instead of median will bias predictions upwards even more for lower risk cases than for higher risk cases (despite the fact, in this case, that the sample size on the left is 1756, compared with the sample size of only 128 on the right; we observe similar default counts for both groups despite the larger sample size for the lower risk case because the sample size difference is offset by higher default risk for the higher risk case).

These results can also be used to demonstrate the impact of asset correlation (non-independence) on the default distribution. Consider the following summaries of the lower and higher risk groups:

Risk	N	P	mean	st.dev.	CV	NP(1-P)
Lower	1756	0.0232	40.74	6.31	15.49	39.7 (>10)
Higher	128	0.20	25.6	4.53	17.70	20.5 (>10)

As discussed before, NP(1-P)>10 here means that the default distribution should be normally distributed, hence symmetric, if defaults are independent. As we can see, NP(1-P) for the lower

---

[38] At first glance this result is somewhat surprising given the small sample size. The result likely relates to the fact, discussed before, that a normal approximation approaches the binomial distribution faster for a given sample size with high risk firms than for low risk firms.

risk group (=39.7) and the higher risk group (=20.5) are both well above the level mentioned earlier (5 -10) for the distributions to be approximately normal (assuming independent defaults). In other words, the skewness we are seeing is driven by the positive asset correlation (0.167 and 0.136 for the lower and higher risk groups, respectively). For comparison, we recreated the previous charts, now based on simulations using zero asset correlation. These results are presented below, again with the lower risk / higher risk groups on the left and right, respectively:
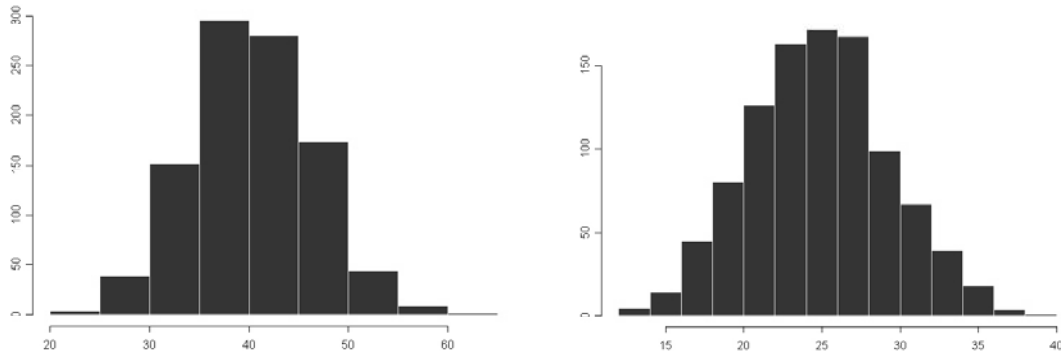
As we can see, changing from the observed positive asset correlation to zero asset correlation removes the skewness entirely.

4.2 RESULTS BY EDF VALUE SUB-GROUPS: CAN WE USEFULLY SLICE THE DATA ANY FURTHER?

The previous results are based on EDF values aggregated into two larger buckets. However, it is sometimes of interest to attempt to validate EDF levels in narrower EDF bands. We now present such results, decomposing the "EDF value < 20" group into three sub-groups. Two statistical points from before that are relevant here are:
- We expect to see more sampling error for smaller samples, so we expect more sampling error for the sub-groups here than for the larger group from which they were taken;
- We expect to see more sampling error for lower EDF value groups.

The three EDF value sub-strata, along with median and mean asset correlations and EDF values for each, are:

Stratum	EDF range	Correl: median/mean	EDF value: median/mean	N[39]
1	0.02 – 5⁻	0.164 / 0.170	0.35 / 0.76	18030
2	5 – 12⁻	0.143 / 0.146	7.13 / 7.57	955
3	12 - 20⁻	0.135 / 0.139	15.11 / 15.51	293

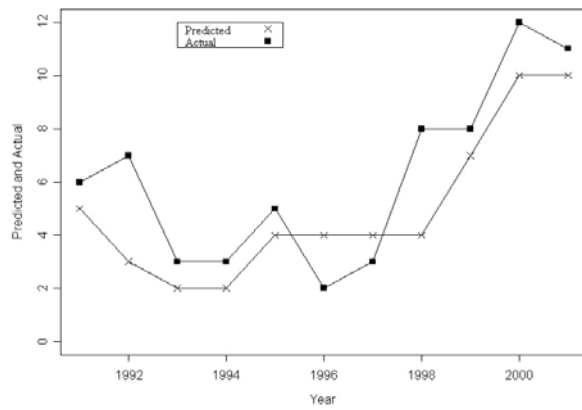We next examine median predicted versus actual defaults over time for each of the three strata.

---

[39] Note that sample sizes by year are of course about an order of magnitude smaller – see appendix.
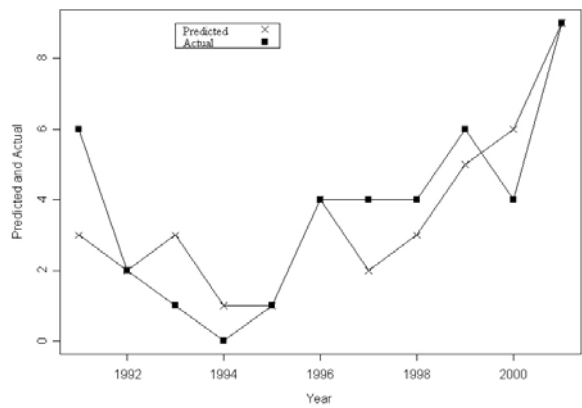
Predicted and actual results for EDF values in the range [0.02, 5):
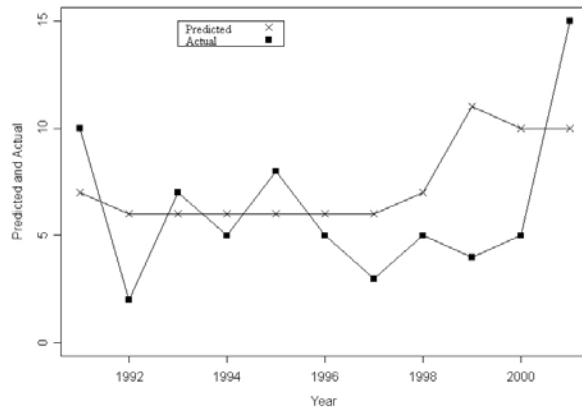
Predicted and actual results for EDF values in the range [5, 12):


Predicted and actual results for EDF values in the range [12, 20):

Comments on the sub-group results:

- Are the results just shown reasonable given sampling error? As a case study, consider the portion of the previous simulation corresponding only to 2001 for EDF values in the range [0.02, 5). In this simulation, although the median number of simulated defaults was 10, 33% of the iterations ended with 5 or fewer defaults, and 38% of the trials ended with 15 or more defaults (15 was the number observed). In other words, the observed results are consistent with predicted results given sampling variability. We expect to see a fair amount of sampling variability in this [0.02, 5) range, as it is the case with the lowest EDF values and highest asset correlations.[40]

- Is there an explanation for the fact that tracking of predicted and actual defaults seems "noisier" for EDF values in the range [0.02, 5) starting around 1997, relative to earlier years? Given the sampling error just discussed it is difficult to draw firm conclusions. That said, it is interesting to compare results there with the yearly volatility of VIX levels shown below.[41] VIX can be viewed as a measure of option-implied systematic risk, and so would be more closely related to the volatility of the large firms in this group than the smaller firms in the higher EDF value groups. The volatility of VIX measures the degree of instability in VIX; note that it increases in the same "noisy" period mentioned above. It seems plausible that the increased uncertainty regarding volatility in this period might correspond to greater sampling variability when comparing predicted and actual defaults. This interpretation is supported by the fact that the rank correlation between the volatility of VIX and the absolute value of median predicted minus actual defaults is +0.48.

Volatility of VIX

---
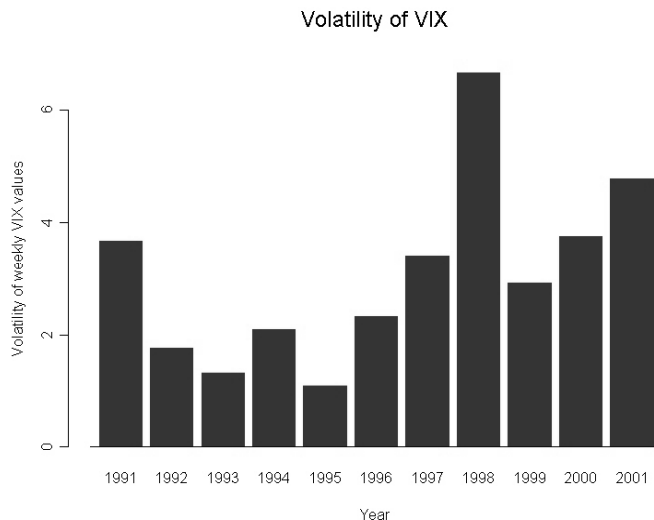
[40] Another way to look at this issue is to consider the CV for the 2001 data in the EDF range [0.02,5) case. Here we have N=1521, P=0.0114, st.dev.=4.14, and CV=24.1. This CV number is substantially larger than the CV number of 15.5 for the 2001 data in the EDF range [0.02,20) cases, so we expect more sampling variability.
[41] VIX represents option-implied volatility on the SP100. To calculate the "vol of vol" numbers used here, we took VIX from the Wednesday of each week (to get a weekly number), then took the standard deviation of the resulting numbers for each year. These are "raw" vol of vol numbers, in that they have not been annualized. VIX data are from www.cboe.com.

In summary, these sub-groups are perhaps too small for reliable validation given that each group contains only large firms, in a particular EDF value range, in a given year. Nonetheless, the results do appear to indicate reasonable estimates of default frequencies, given sampling error.

5. Some International Results Using Rated Firms

Another way to choose a subset of data in which hidden defaults are less likely to be an issue is to choose rated firms. Here we illustrate this approach on international firms that have Moody's ratings.[42] We again compare median predicted default counts from simulation with observed defaults.[43] The results again seem reasonable given the small samples and resulting sampling error:
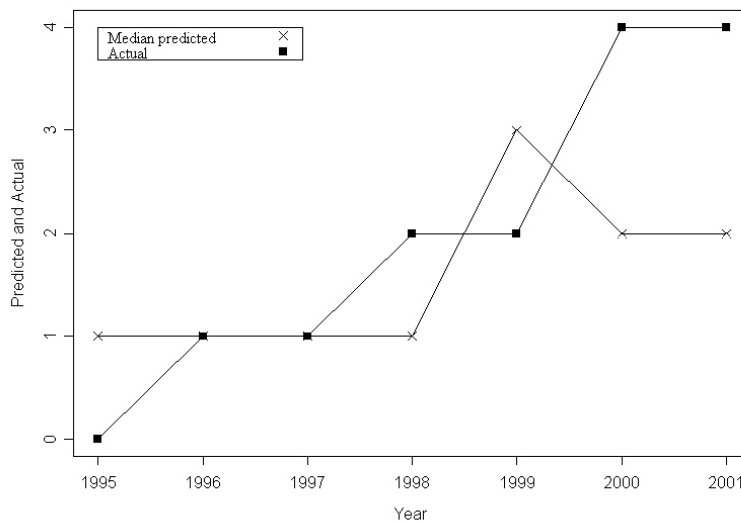
Here are the data underlying the previous graph:

```
Year   Firms Predicted      Defaults
1995    154   1                    0
1996    172   1                    1
1997    175   1                    1
1998    187   1                    2
1999    186   3                    2
2000    180   2                    4
2001    175   2                    4
Total  1229  11                   14
```

---

[42] Rated defaults included here may not match those supplied by Moody's: ratings are from EJV Bridge, and EJV's list of rated firms does not exactly match Moody's list.
[43] The median asset correlation for this set of firms is 0.11.

6. Other Issues in Level Validation

Here we briefly discuss several relevant issues.

6.1 DEFAULT DATA

Level validation is more challenging for smaller firms than for larger firms because default information on smaller firms is less readily available. For example, large firm defaults are reported more often in the news sources that are often used to populate default data sets. One clue that this may be a problem in a database is if the ratio of defaults to bankruptcies is much lower for small firms than for large firms.[44]

Also, some data sets are not designed for research purposes. If, for example, a lender tends to drop internal ratings/data on firms just prior to default, then research using such a database would tend to underestimate default risk.

6.2 HOW DOES SKEWNESS IMPACT MODEL CALIBRATION?

Does the skewness of mean default rate distributions negatively impact attempts to *calibrate* (as opposed to validate) a model? If defaults and default predictions span a long period of time, such as one or more decades,[45] then the answer is essentially "no." This is because, while asset correlations are positive within shorter time spans such as a year, changes in asset value of two different firms at two different points in time should become uncorrelated as the two points become more separated in time. Under these circumstances, differences in mean and median predicted default rates become small relative to sampling error.[46]

6.3 MODEL COMPARISON ISSUES[47]

The results shown previously have no implications for testing of model *power*, as opposed to *calibration*, because power tests are based solely on rank orders of model scores and not absolute levels. Nevertheless, there is at least one interesting relationship between power and calibration, as we see presently. The issue we address here is tangential to the main points of this paper, but we mention it because it is important.

Sometimes two models are compared against each other on a sub-population of firms, and the average predicted level of defaults is higher for one model than for the other. This situation can result in an interpretation that the *level* of one or the other must be wrong, because they cannot both be right.

Differences in power between models may lead to differences in apparent calibration for particular sub-groups of firms, even for models that predict the same overall level of default risk for the entire population of firms. For example, say "model" A simply assumes a historical average default rate of 2%, while model B is a more powerful model but still predicts an overall default rate of 2%. Say we select a sub-group of firms that does not have the average default risk

---

[44] We at Moody's | KMV are actively researching this issue of "hidden" defaults.
[45] This is the case with Credit Monitor and CreditEdge.
[46] Cantor et al. (2001) present a comprehensive discussion of persistence in default rate shocks over time.
[47] I thank Roger Stein for bringing this issue to my attention.

for the population – for example, firms from the smallest 25% of the population (smaller firms have higher than average default risk). Model A will still predict a 2% default rate for these smaller firms, while Model B will predict a default rate above 2% (assuming B's predictions are correlated with size) because the firms involved have above-average default risk. In this case, the two models will spuriously appear to be mis-calibrated relative to each other, when in fact the difference in performance has nothing to do with calibration but with power. The same point applies even if model A is more powerful than the simple mean default rate, so long as models A and B differ in power, and we select a subset of firms from the larger set based on a characteristic (like size) that is correlated with model predictions.

The important point to note here is that this sort of situation can result from differences in power between two models, even if the two models have the same average predicted default risk level across the entire population. To put the point differently, what may appear to be a mis-calibration in levels for one or the other model may actually be driven by differences in power between the two. (For more discussion of this point see Stein, 2002).

7. Summary

Asset returns across firms are correlated, leading to skewed distributions of predicted numbers of defaults in a portfolio. Such skewness is higher for groups of firms with higher asset correlation, lower default risk, and smaller samples sizes. This skewness in turn implies that mean default rates should exceed median default rates.

A "fair" comparison between predicted and actual default rates must account for correlation in particular and skewness more generally. This fair comparison can be performed by first simulating a default distribution, and then using the median of the simulated distribution to obtain the predicted default rate. Skewness issues are of major importance in annual comparison tests between predicted and actual defaults, but are of minor importance when the predicted and actual defaults are aggregated over one or more decades.

Ignoring skewness in default distributions will tend to *erroneously* make predicted default rates *seem* too high, and that this spurious bias will be greatest for large, low risk (e.g., investment grade) firms.

The degree of sampling error in default rates is an important, related issue. For fixed sample sizes, we can increase the precision of default rate estimates by increasing the range of EDF values used in a given sample, or the length of the time period over which they are collected. Grouping firms in this way reduces sampling error, but at the same time reduces the number of EDF-value-subgroups over which we can check the level of the measure.

Testing over the time period 1991-2001 demonstrates that the EDF credit measure is a good predictor of actual defaults, both in terms of their ability to predict the overall level of defaults (calibration), and in their ability to discriminate defaults from non-defaults (power).

8. References

Cantor, Richard, Ted Collins, Eric Falkenstein, David Hamilton, Chee Mee Hu, Chester Murry, Sameer Nayar, Robert Ray, Everett Rutan, and Farisa Zarin (2001). Testing for Rating Consistency in Annual Default Rates. Moody's Investors Service.

Crosbie, Peter and Jeff Bohn (2001). Modeling Default Risk. Moody's | KMV Technical Document.

Das, Ashish, and Shota Ishii (2001). Understanding Correlation Differences. Moody's | KMV Technical Document.

DeFusco, Richard, Dennis McLeavey, Jerald Pinto and David Runkle (2001). Quantitative Methods for Investment Analysis. Baltimore: United Book Press.

Evans, Merran, Nicholas Hastings and Brian Peacock (1993). Statistical Distributions (2nd Ed.).

Hamilton, David, Richard Cantor and Sharon Ou (2002). Default and Recovery Rates of Corporate Bond Issuers. Moody's Investor's Services.

Johnson, Norman, Samuel Kotz, and Adrienne Kemp (1993). Univariate Discrete Distributions, 2nd Ed., NY: Wiley.

Kealhofer, Stephen (2002a). Quantifying Credit Risk, Part I Default Prediction. Paper in press, *Financial Analysts Journal.*

Kealhofer, Stephen (2002b). Quantifying Credit Risk, Part II Debt Valuation. Paper in press, *Financial Analysts Journal.*

Kealhofer, Stephen, and Jeff Bohn (2001). Portfolio Management of Default Risk. Moody's | KMV Technical Document.

Kealhofer, Stephen, Sherry Kwok and Wenlong Weng (1998). Uses and Abuses of Bond Default Rates. Moody's | KMV Technical Document.

Stein, Roger (2002). Benchmarking Default Prediction Models: Pitfalls and Remedies in Model Validation. Moody's Risk Management Services, Technical Report #020305.

Stuart, Alan and Keith Ord (1994). Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory (6th Ed.)

Vasicek, Oldrich, (1991). Limiting Loan Loss Probability Distribution. Moody's | KMV Technical Document.

Zeng, Bin, and Jing Zhang (2001). An Empirical Assessment of Asset Correlation Models. Moody's | KMV Technical Document.

9. Appendix

This appendix has the following parts:
- A review of the procedure for simulating correlated defaults.
- Results showing that the validation work presented here is not overly sensitive to the values of asset correlation used.
- A summary of the year-by-year data from the section on results by risk group.

9.1 SIMULATION OF CORRELATED DEFAULTS

We perform simulation of correlated defaults by simulating correlated asset value changes, then converting those to correlated defaults.[48] We discuss each of these things below.

9.1.1 A simplified approach

As explained in the body of the paper, we use a single median asset correlation to perform simulations because it is a reasonable approximation. Later in the appendix, we examine sensitivity of our results to the value of correlation used.

The actual Moody's | KMV Gcorr approach to modeling correlation is through a multi-factor model, but we only need a single factor model here because that suffices if we use a single median correlation. For those who are unwilling or unable to use something like Gcorr to perform their own validation work, this approach is a good alternative to ignoring the effects of correlation entirely.

9.1.2 Why use asset correlations instead of equity correlations?

Equity correlations are not a good proxy for asset correlations. Two reasons for this are: (1) equity correlations are less stable (less good as forecasts of future correlation) because they mix in changes in leverage with changes in firm value; (2) equity correlations tend to underestimate asset correlations. For more details, see Zeng and Zhang (2001).

9.1.3 Simulation of multivariate normal asset returns with a single constant correlation between all firms

The asset value distribution is modeled as multivariate normal.[49] In the stochastic simulation literature, techniques such as Cholesky Decomposition and Eigen-Decomposition are typically used to simulate multivariate normal draws. However, such techniques do not scale well to larger numbers of correlated variables, so here we use a factor model for simulation.

Consider two firms whose systematic risks are explained by the market portfolio, R_m. Mathematically,

$$R_j = \beta_{jm}R_m + \varepsilon_j \qquad \text{and}$$
$$R_k = \beta_{km}R_m + \varepsilon_k$$

Where

β = sensitivity of an asset's return to variations in the market portfolio

ε = idiosyncratic risk (assume zero mean returns for all series over short horizons)

---

[48] For more details, see Kealhofer and Bohn (2001).

[49] For those who would advocate a more fat-tailed approach to modeling the joint distribution, note that while it is straightforward to observe non-normality in *observed* return series, it is much more difficult to *forecast* it with any notable degree of accuracy over the 1-5 year time horizons over which we forecast EDF values.

In this case, individual firm variances are:

$$\sigma_j^2 = \beta_{jm}^2 \sigma_m^2 + \sigma_{\varepsilon j}^2$$
$$\sigma_k^2 = \beta_{km}^2 \sigma_m^2 + \sigma_{\varepsilon k}^2$$

The covariance between a pair of firms is:[50]

$$
\begin{aligned}
\mathrm{cov}(R_j, R_k) \quad &= E[R_j R_k] \\
&= E[(\beta_{jm} R_m + \varepsilon_j)(\beta_{km} R_m + \varepsilon_k)] \\
&= \beta_{jm}\beta_{km} E[R_m^2] + E[\varepsilon_j \varepsilon_k] \\
&= \beta_{jm}\beta_{km} \sigma_m^2
\end{aligned}
$$

Consequently, the correlation is:

$$\rho_{jk} = \mathrm{cov}(R_j, R_k)/\sigma_j \sigma_k$$
$$\rho_{jk} = \beta_{jm}\beta_{km}\sigma_m^2 / \{(\beta_{jm}^2 \sigma_m^2 + \sigma_{\varepsilon j}^2)(\beta_{km}^2 \sigma_m^2 + \sigma_{\varepsilon k}^2)\}^{\frac{1}{2}}$$

If we fix correlation, beta, and idiosyncratic risk to be constant across firms, this becomes:

$$\rho = \beta^2 \sigma_m^2 / \{(\beta^2 \sigma_m^2 + \sigma_\varepsilon^2)(\beta^2 \sigma_m^2 + \sigma_\varepsilon^2)\}^{\frac{1}{2}}$$

If we also fix $\sigma_m^2 = \sigma_\varepsilon^2 = 1$ for simplicity (standard normal iid draws for our market and idiosyncratic risk factors), this becomes

$$\rho = \beta^2/(\beta^2 + 1)$$

or

$$\beta = \{\rho/(1-\rho)\}^{\frac{1}{2}}$$

In this construction, variance of the returns of asset j equals:

$$\mathrm{var}(R_j) = \beta_{jm}^2 \mathrm{var}(R_m) + \mathrm{var}(\varepsilon_j) = \beta_{jm}^2 + 1$$

so

$$\sigma_{Rj} = (\beta_{jm}^2 + 1)^{\frac{1}{2}}$$

So, we divide each asset return by its standard deviation to make it standard normal. Summarizing, we have each simulated standard normal asset return R_j^n as the following function of correlation and iid $N(0,1)$ market (R_m) and idiosyncratic (ε_j) returns:

$$R_j^n(\rho, R_m, \varepsilon_j) = (\beta R_m + \varepsilon_j)/\sigma_{Rj}$$
$$\beta = \{\rho/(1-\rho)\}^{\frac{1}{2}}$$
$$\sigma_{Rj} = (\beta^2 + 1)^{\frac{1}{2}}$$

To apply these results, draws of the market factor should of course be the same across firms for any given iteration.

9.1.4 Conversion of correlated asset returns to correlated defaults

Each asset's return distribution is modeled as a standard normal. We treat each simulated draw R as a default if and only if $R < N^{-1}(\mathrm{EDF})$, where N^{-1} is the inverse cumulative normal function, and EDF here is scaled to be a probability (0-1) instead of a percentage.[51]

---

[50] $E[R_m^2] = \sigma_m^2$ because R_m^2 is chi-squared distributed with df=1; $E[\varepsilon_j \varepsilon_k]=0$ and given ε_j and ε_k iid normal.
[51] Kealhofer and Bohn (2001) discuss the joint default correlation that results from using this approach.

9.1.5 Validation of these results via simulation

As a quality control check, we simulated two firms asset returns (Rj and Rk) for various levels of population correlation (0.0, 0.1, 0.2, …, 0.9) between the firms. We simulated a pair of firms 10,000 times at each level of correlation, and found that:

- The correlation between predicted and actual correlations across these iterations is 0.9999.
- The intercept and slope of the regression line fitting predicted and actual correlations are –0.002 and 1.006, respectively (0 and 1 would indicate perfect fit).
- A graph of predicted versus actual correlations demonstrates the relation between the two to be linear.

9.2 HOW SENSITIVE ARE OUR VALIDATION ASSUMPTIONS TO THE VALUE USED FOR CORRELATION?

The results presented in this paper are essentially a joint test that both EDF values and asset correlations are well modeled. To decouple these two pieces to some degree, we repeated the simulations done for firms with EDF values less than 20, this time using 0.1 and 0.2 as the single asset correlation (recall that the asset correlation value used for this group of firms was 0.167). The results, presented below, show that

- The median predicted default counts begin to look slightly low on average for correlation of 0.2;
- The median predicted default counts begin to look somewhat high on average for correlation of 0.1.

As noted previously, the mean default rate is what we obtain using correlation of 0.0, and predictions there looked too high on average. In combination, these results provide some additional support for the claim that asset correlation is needed for level validation.

Results for asset correlation of 0.1:
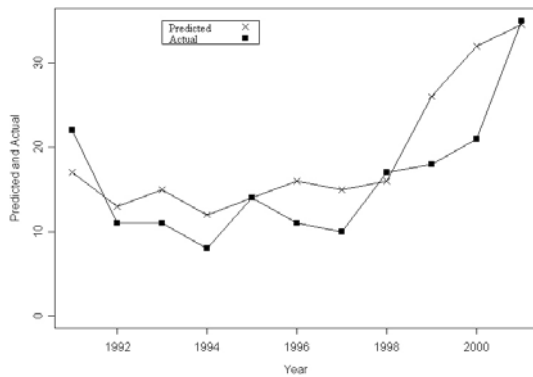
Results for asset correlation of 0.2:
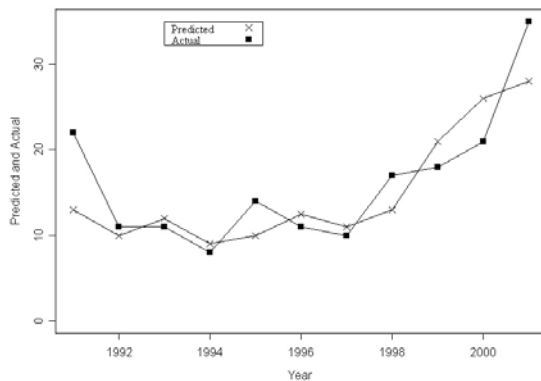
9.3 SUMMARY OF DATA UNDERLYING FIGURES IN SECTION ON RESULTS BY RISK GROUP

Here we summarize the year-by-year data from the section on results by risk group.

Stratum #1:

Year	N	MeanPredicted	MedianPredicted	Defaults
1991	1351	11.30	7	10
1992	1416	9.32	6	2
1993	1505	10.92	6	7
1994	1611	10.91	6	5
1995	1624	10.98	6	8
1996	1724	10.41	6	5
1997	1873	10.53	6	3
1998	1938	12.63	7	5
1999	1878	17.96	11	4
2000	1589	16.02	10	5
2001	1521	17.18	10	15

Stratum #2:

Year	N	MeanPredicted	MedianPredicted	Defaults
1991	83	6.41	5	6
1992	52	4.23	3	7
1993	45	3.19	2	3
1994	44	3.27	2	3
1995	73	5.43	4	5
1996	65	4.64	4	2
1997	62	4.80	4	3
1998	69	5.50	4	8
1999	113	8.50	7	8
2000	180	13.26	10	12
2001	169	12.74	10	11

Stratum #3:

Year	N	MeanPredicted	MedianPredicted	Defaults
1991	23	3.52	3	6
1992	14	2.15	2	2
1993	24	3.72	3	1
1994	12	1.82	1	0
1995	9	1.25	1	1
1996	27	4.31	4	4
1997	18	2.56	2	4
1998	21	3.60	3	4
1999	36	5.66	5	6
2000	43	6.56	6	4
2001	66	10.08	9	9