Catching Fallen Angels (and Other Expensive Credit Events)

Lisa R. Goldberg*
MSCI Barra, Inc.

Andreas Zapp†
Bundesanstalt für Finanzdienstleistungsaufsicht

September 13, 2005‡

1 Introduction

Credit risk is the distribution of loss due to change in the credit quality of a financial counterparty. Default risk, or probability of default over a time horizon, is one of the main drivers of credit risk. Credit models are often evaluated solely on their ability to forecast default. In this article, we examine the efficacy of the I^2 credit model, which is the engine of BarraCredit, in a broader context that is relevant to fund and asset managers.

Using a rigorous statistical analysis, we show that I^2 is a powerful forecaster of the following events:

- Rating agency downgrades
- Investment grade to high yield downgrades
- High yield defaults

Early identification of investment grade firms that are downgraded to high yield, also known as *fallen angels*, is central for portfolio managers who are subject to investment policy restrictions. High yield defaults are important to managers invested in very speculative names.

*MSCI Barra, Inc., 2100 Milvia Street, Berkeley, CA 94704-1113, USA, Phone (510) 649 4601, Fax (510) 848 0954, email: `lisa.goldberg@mscibarra.com`.

†Graurheindorfer Str. 108, 53117 Bonn, Germany, `andreas.zapp@bafin.de`. The statements made in this article represent the author's view and do not necessarily reflect the view of the BaFin.

‡We are very grateful to Kay Giesecke, Rajnish Kamat and Vijay Poduri for their contributions to this article.

Our statistical results lead to effective portfolio management and construction strategies, which are described in Section 7. For example, the probability of a portfolio experiencing an expensive credit event can be reduced significantly by selectively pruning the universe of investable names.

The paper is organized as follows. Section 2 sketches model details and Section 3 contains a description of the statistical methods used in our analysis. In Section 5 we give the details of our data set and we present our analysis of the discriminatory power of I^2 in Section 6. Finally, we evaluate different portfolio strategies based upon the previous results.

2 The I^2 Model and BarraCredit

The I^2 model is a hybrid that reflects the two prevailing approaches in credit risk modelling, the structural approach and the reduced form approach. The coherent combination of these approaches stems from the assumption of *incomplete information*: investors are not perfectly informed when they execute trades. Background on incomplete information models can be find in (Giesecke 2001). The specification of I^2 as well as empirical examples of its output are in (Giesecke & Goldberg 2004).

BarraCredit provides daily I^2-based martingale default probabilities (BDPs) for nearly 10,000 firms in the US and Europe. In (Poduri 2004), it is shown that I^2 is a power forecaster of default and that its discriminatory power trumps agency ratings.

3 Power Curves and Accuracy Ratios

Default probability forecasts are used to rate firms by credit quality. The Basel II Accord calls for validation of this type of rating system, see (Basel Comittee on Banking Supervision 2005). A standard validation tool is an *ROC* or *power* curve.

To understand how a power curve works, suppose we have a rating R on a universe F of firms. For simplicity, we assume R is continuous, and this assumption is satisfied by the ratings we consider below. Divide F into a pair of disjoint subsets. The set F^+ consists of rated firms for which a specified event, such as a downgrade or default within one year from the time the rating is issued, occurs. The non-event set F^- is its complement. Fix a cutoff c for the default probabilities. In a discriminating model, the fraction of *true positives*

$$\mathrm{TP}(c) = P(R > c \,|\, F^+)$$

should be relatively large and the fraction of *false positives*

$$\mathrm{FP}(c) = P(R > c \,|\, F^-)$$

should be relatively small. A power curve is a parametric plot of true positives against false positives as a function of cutoff. A diagonal power curve corresponds to a coin toss: the ratings are indiscriminate with respect to event under consideration. A curve that lies far above the diagonal corresponds to a model with good discriminatory power.

With some loss of information, a power curve can be condensed into a single number. The *accuracy ratio* of a power curve is the area it bounds. The accuracy ratio takes values in $[0, 1]$ and a higher ratio indicates greater discriminatory power. A schematic illustration of a power curve and its accuracy ratio are in Figure 1.

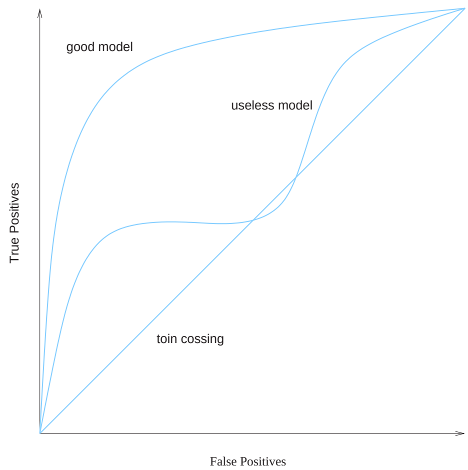

Figure 1: Sketches of power curves. The diagonal corresponds to a coin toss, the curve in the upper left represents a good model. The third curve represents a useless model despite its high accuracy ratio.

The accuracy ratio can be estimated with the *Mann-Whitney* U-statistic, which is defined in terms of a pair of ranked populations.[1] It is an unbiased estimator of the probability that the rank of a random draw from the first population is exceeded by the rank of a random draw from the second. In the examples below, it estimates the probability that the BDP of a firm that does not experience a credit event is less than the BDP of a firm that does.[2] For each pair of rated firms $(f^+, f^-) \in F^+ \times F^-$, set

$$u(f^+, f^-) = \begin{cases} 1 & \text{if } R(f^+) > R(f^-) \\ 0 & \text{otherwise} \end{cases}$$

and

$$\hat{U} = \frac{1}{|F^+ \times F^-|} \sum u(f^+, f^-)$$

[1]A general reference on estimators of the Mann-Whitney statistic is (Lehmann 1998).

[2]See (Bamber 1975) and (Engelmann, Hayden & Tasche 2003), who introduced this statistic to the credit risk literature.

where the sum is taken with respect to all pairs $f^+ \in F^+$and $f^- \in F^-$ and $|\cdot|$ indicates cardinality. The relationship between the accuracy ratio and the Mann Whitney statistic comes through the formula for the area under a parametric curve:

$$\begin{aligned}
\mathrm{AUC} &= -\int_0^1 \mathrm{TP}(c)\, d\,\mathrm{FP}(c) \\
&= -\int_0^1 P(R > c \,|\, F^+)\, dP(R > c \,|\, F^-) \\
&= \int_0^1 P(R > c \,|\, F^+)\, dP(R \leq c \,| F^-) \\
&= P(R(f^+) > R(f^-))
\end{aligned}$$

where f^+ and f^- are independent draws from the defaulting and non-defaulting sets F^+ and F^-. It follows immediately that

$$\mathrm{AUC} = E(\hat{U})\,.$$

Since the variance $\sigma^2_{\hat{U}}$ is finite, a standard argument based on the central limit theorem implies that the distribution of the standardized Mann-Whitney statistic

$$\frac{\hat{U} - \mathrm{AUC}}{\hat{\sigma}_{\hat{U}}}$$

is asymptotically distributed as a standard normal. In the applications below, a bootstrapping method is used to verify the approximate normality $\hat{U}$. Hence, an α-confidence interval for AUC is roughly

$$\left[\hat{U} - \hat{\sigma}_{\hat{U}}\Phi^{-1}\left(\frac{\alpha+1}{2}\right), \hat{U} + \hat{\sigma}_{\hat{U}}\Phi^{-1}\left(\frac{\alpha+1}{2}\right)\right]$$

where Φ denotes the cumulative distribution function of the standard normal.

The Mann-Whitney statistic can be used to statistically distinguish ratings R^1 and R^2 on a universe F. Following (Engelmann et al. 2003), we consider the test statistic

$$T = \frac{(\hat{U}^1 - \hat{U}^2)^2}{\hat{\sigma}^2_{\hat{U}^1} + \hat{\sigma}^2_{\hat{U}^2} - \hat{\sigma}^2_{\hat{U}^1,\hat{U}^2}}$$

where $\hat{\sigma}^2_{\hat{U}^1,\hat{U}^2}$ denotes the covariance between the estimators $\hat{U}^1$ and $\hat{U}^2$. A formula for the covariance and a proof that T is asymptotically distributed as $\chi^2(1)$ is in (DeLong, DeLong & Clarke-Pearson 1988).

4 Application of Power Curves to Portfolio Management

Consider a portfolio manager who invests in a universe of corporate bonds, credit default swaps, or other credit sensitive securities. Suppose the securities in the universe are ranked by creditworthiness, with the worst securities corresponding to the largest rankings. A simple strategy is to exclude the securities with the largest ranks before constructing a portfolio. How effective is this strategy at helping the manager avoid a downgrade or default? More precisely,

1. By how much is the probability of default or downgrade reduced when the universe of investment opportunities is truncated by a definite fraction consisting of securities with the worst rankings?

2. What is the optimal fraction of securities to exclude?

Bayes' theorem provides a theoretical framework in which to answer the first question. Let $p = P(F^+)$ denote the (unconditional) probability of an event occurring and let $\gamma \in (0,1)$ denote the fraction of the universe that the manager is willing to exclude. From the power curve, we can find the cutoff value $c = c(\gamma)$ such that the probability $P(R > c)$ that the ranking exceeds c is equal to γ. We are interested in the conditional probability of a credit event on the residual universe of firms for which $R \leq c$, which is given by

$$P(F^+ \mid R \leq c) = \frac{P(R \leq c \mid F^+)\, P(F^+)}{P(R \leq c)} \tag{1}$$

$$= \frac{(1 - \mathrm{TP}(c))\, p}{1 - \mathrm{TP}(c)p - \mathrm{FP}(c)(1-p)}. \tag{2}$$

When the cutoff c is at its maximum, then $\mathrm{FP}(c) = \mathrm{TP}(c) = 0$ and the conditional probability $P(F^+ \mid R \leq c)$ is equal to the unconditional probability p. If the ranking R provides discriminating event forecasts, then the associated power curve is initially very steep and an auspicious choice of γ results in a large power ratio $\mathrm{TP}(c)/\,\mathrm{FP}(c)$. For rare events such as defaults or downgrades, the unconditional probability p is already quite small. From formula (2), it follows that the conditional probability $P(F^+ \mid R \leq c)$ is much less than p in this situation.

To answer the second question, suppose the portfolio manager is willing to exclude at most a fixed fraction γ^* of the universe. We consider the constrained optimization problem

$$\min_c P(F^+ \mid R \leq c) \quad \text{with}\, c > c(\gamma^*).$$

If the power curve is concave, this problem has a unique solution given by the critical value $c = c(\gamma^*)$. The assumption of concavity is a reasonable one, and it is satisfied empirically by power curves below.

5 Data

The data set is based on roughly 1500 S&P rated firms. For each firm, we have a one year default default probability forecast (BDP) for each of the 912 business days during the period beginning December 2001 and ending July 2005. The distribution of ratings in mid April 2005 is the following

S&P Rating	BDP	*pct.*
AAA	12	0.8%
AA	41	2.8%
A	258	17.7%
BBB	486	33.3%
BB	388	26.6%
B	245	16.8%
CCC	24	1.6%
CC	2	0.1%
D	5	0.3%
total	1461	100.0%
Investment Grade	797	54.6%
High Yield	664	45.4%

Table 1: Most firms have a ratings in the range from single A to single B.

The following events are covered by the data set, see Table 2. Note that a downgrade by a single notch is counted as an event.

Event	Number	annualized Event probability p
Downgrade (no default)	974	20.3%
Downgrade from BBB to HY	106	4.0%
Upgrade	416	7.8%
Default	51	1.1%
Default of HY names	49	1.8%

Table 2: Events are relatively rare.

6 Results

We use the Barra default probabilities described in Section 2 to rank firms and we use power curves to evaluate their efficacy at forecasting credit events.

In our analysis, we consider the *timeliness* of the information in the rating. Every choice of a time horizon between the BDP forecast and the event in question generates a different power curve. As explained in Section 3, these power curves can be compared by means of the test statistic T on the AUCs. We can decide if there is a significant difference between the information provided by BDPs at different time horizons. In the results below we assume approximate normality of the Mann-Whitney statistic. Bootstrap tests demonstrating the validity of this assumption are in Appendix A.

6.1 Downgrades

Figure 2 illustrates the strong discriminatory power of the Barra default probability with respect to agency downgrades. Accuracy ratio estimates and confidence intervals are in Table 3.

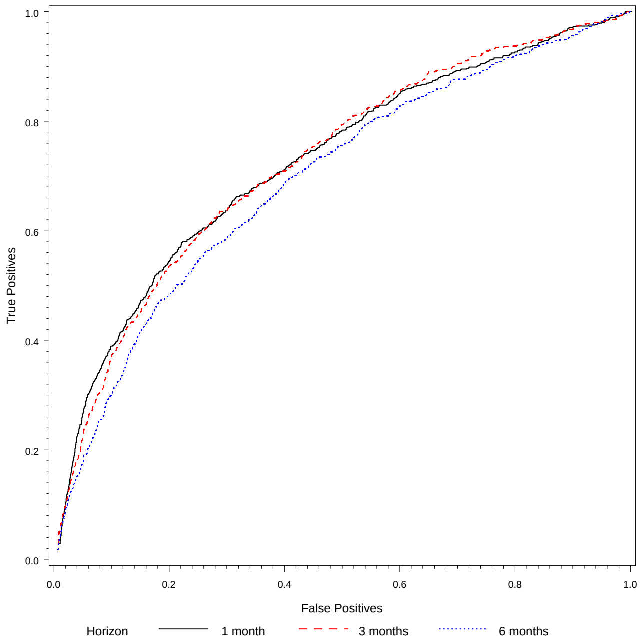

Figure 2: Power curves for the I^2 model with respect to downgrades. The three curves correspond 1, 3 and 6 month time horizons between the forecast and the event. All three curves indicate that BDP forecasts have good discriminatory power with respect to downgrades.

AUC and 99% Confidence Intervals				
Horizon	**AUC**	**Std Error**	**Lower**	**Upper**
1 month	0.7265	0.0099	0.7010	0.7519
3 month	0.7247	0.0095	0.7002	0.7493
6 month	0.6935	0.0099	0.6679	0.7191

Table 3: Accuracy ratio estimates and confidence intervals for power curve evaluations of BDP forecasts of downgrades at three time horizons.

The results in Table 4 show that T-statistic does not distinguish the performance of the 1 month and the 3 month forecasts. Both are superior to the 6 month forecast.

	T	p-value
1m vs. 3m	0.2018	0.6533
1m vs. 6m	27.7639	<.0001
3m vs. 6m	34.5480	<.0001

Table 4: Test for the null hypothesis that distinct power curves have the same accuracy ratio: The T-statistic is asymptotically $\chi^2(1)$ distributed. The high $p - value$ for the 1 month against 3 month forecast indicates that one cannot reject the null hypothesis.

6.2 Fallen Angels

How effective are BDPs at identifying fallen angels?[3] The relevant power curves are in Figure 3 and accuracy ratio are below in Table 5.

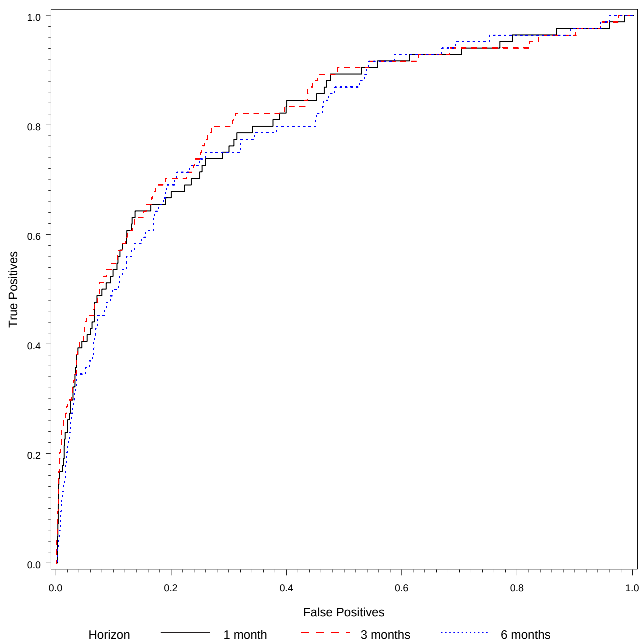

Figure 3: Power curves for the I^2 model with respect to fallen angels at different time horizons. All three show that BDP forecasts have good power to discriminate downgrades.

[3]Our data set contains no direct migrations from A or better to the high yield universe. Hence all fallen angels are downgrades from BBB.

AUC and 99% Confidence Intervals				
Horizon	**AUC**	**Std Error**	**Lower**	**Upper**
1 month	0.8111	0.0264	0.7430	0.8792
3 month	0.8192	0.0264	0.7511	0.8872
6 month	0.7998	0.0261	0.7326	0.8670

Table 5: Accuracy ratio estimates for power curve evaluations of BDP forecasts of fallen angels at different time horizons.

The power curves show excellent detection of fallen angels by the Barra default probabilities. Accuracy ratios here are higher overall than for power curves corresponding to general downgrades. Furthermore, forecasts at 1, 3 and 6 month horizons are indistinguishable by the T- statistic.

	T	**p-value**
1m vs. 3m	0.5191	0.4712
1m vs. 6m	0.3148	0.5747
3m vs. 6m	0.9923	0.3369

Table 6: No instance of the null hypothesis can be rejected.

6.3 Default of high yield issuers

Finally, we examine the power of the I^2 model to detect defaults within the high yield universe. The power curve shows excellent discrimination. The 1 and 3 month forecasts are better than the 6 month forecast, which nevertheless has substantial power. Results are summarized in Figure 4 and Tables 7 and 8.

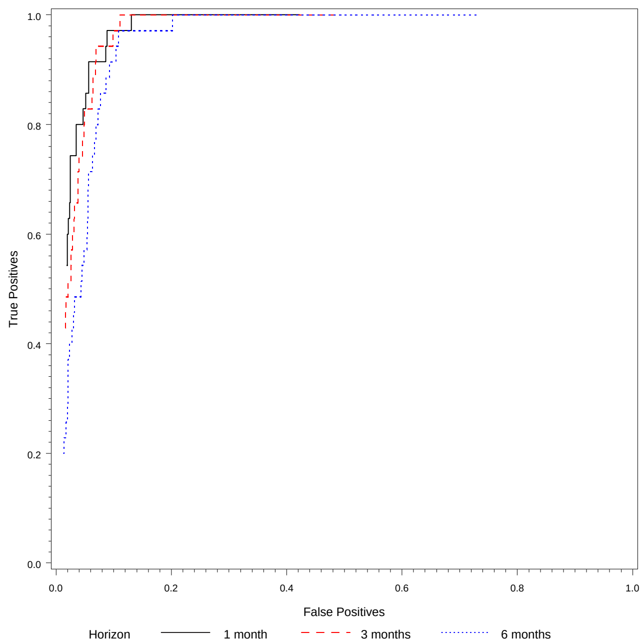

Figure 4: Power curves for the I^2 model with respect to high yield defaults at different time horizons. The model has excellent power at all three horizons.

ROC Curve Areas and 99% Confidence Intervals				
	AUC	**Std Error**	**Lower**	**Upper**
1 month	0.9735	0.0048	0.9612	0.9859
3 month	0.9700	0.0047	0.9579	0.9821
6 month	0.9537	0.0069	0.9359	0.9716

Table 7: Accuracy ratio estimates confirm the visual impression of excellent discriminatory power of the I^2 model with respect to high yield defaults.

	T	**p-value**
1m vs. 3m	1.0208	0.3123
1m vs. 6m	8.8603	$<.0029$
3m vs. 6m	9.3931	$<.0022$

Table 8: The T-statistic test does not distinguish between forecasts at the 1 and 3 month horizon.

7 Management of event risk

Consider a portfolio manager who is restricted to the investment grade universe. The results in Section 6 indicate that he can profitably use the I^2 model to augment his fundamental research. Since BDP forecasts identify names that are likely to be downgraded in the near term, the manager can rebalance his portfolio in a timely fashion. The *price* for this early warning system is a reduction of the allowed universe of names by a specific amount.

We consider the BBB and high yield universes described above. The ordinal rankings imparted by BDP forecasts at a 1 month time horizon are used to exclude names each month. We observe the historical downgrades in the remaining universe of credits compared to the unrestricted case and conclude that a broad majority of events would have been avoided at moderate exclusion rates. The results are in Table 9.

Exclusion rate	Remaining events
0%	106
10%	51
15%	43
20%	34
30%	26

Table 9: The number of fallen angels in the investment universe can be significantly reduced at moderate exclusion rates.

Analogously, a high yield portfolio manager may want to monitor the default risk. As above, we filter the worst high yield names on the basis of the ranking imparted by BDP forecasts at 1 month horizon. At the price of excluding only 10% of the available names, the manager would have reduced his exposure from 49 to 2 defaults during the period December 2001 – July 2005.

Exclusion rate	Remaining events
0%	49
5%	10
10%	2

Table 10: The number of high yield defaults can be reduced dramatically at very low exclusion rates.

A Bootstrapping

Bootstrapping is a general computer-intensive technique used to infer the value of a parameter of interest, see (Efron & Tibshirani 1994) for details. Given a sample X of a population, the bootstrap uses the data in X to construct an estimator $\hat{\theta}$ of the interesting parameter θ. The main idea is to identify the sample X with the population from which X is drawn. Random subsamples X^i are drawn (with replacement) from X and each subsample is used to generate an estimator $\hat{\theta}^i$. The bootstrap estimator $\hat{\theta}$ is the average of the $\hat{\theta}^i$s.

As an example, consider the Mann-Whitney statistic U discussed in Section 3. The sample space consists of pairs, $X = F^+ \times F^-$ each of which consists of an event and a non-event. To find a bootstrap estimator of U, follow these steps:

1. Draw (with replacement) a subsample X^i of X.
2. Estimate the Mann-Whitney statistic $\hat{U}^i$ on each X^i.
3. Use the sample U^is do estimate the quantiles of $\hat{U}$.

In our study, we have used the bootstrap to check if the assumption of approximate normality of $\hat{U}$ is satisfied and subsequently to compute confidence intervals for AUC.

In Table 11 we present the bootstrap results for fallen angels, see section 6.2.

AUC and 99% Confidence Intervals				
Horizon	**AUC**	**Std Error**	**Lower**	**Upper**
1 month	0.8116	0.0264	0.7436	0.8742
3 month	0.8195	0.0264	0.7582	0.8822
6 month	0.8003	0.0246	0.7331	0.8600

Table 11: Bootstrapped accuracy ratios estimates for BDP detecting fallen angels at different time horizons.

The differences between the estimates and the confidence intervals are small, as can be seen by comparing the values in Tables 5 and 11. The distribution of $\hat{U}$ is close to normal as illustrated in Figure 5.

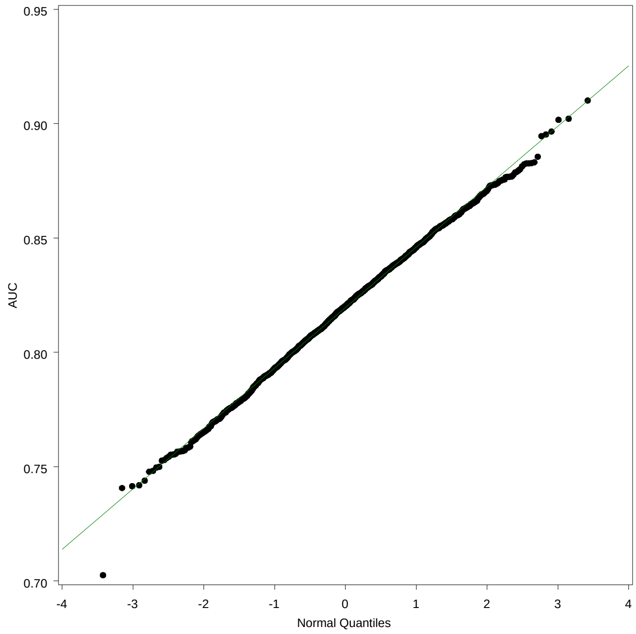

Figure 5: Quantile-Quantile Plot of bootstrapped distribution against fitted normal distribution (3 month forecasts for fallen angels).

Qualitatively similar results were obtained for the two other types of events considered in this articles.

References

Bamber, D. (1975), 'The area above the ordinal dominance graph and the area below the receiver operating graph', *Journal of mathematical psychology* **12**, 387–415.

Basel Comittee on Banking Supervision (2005), 'Studies on the validation of internal rating systems', *Working Paper* (14).

DeLong, DeLong & Clarke-Pearson (1988), 'Comparing the area under two or more correlated receiver operating characteristic curves: a nonparametric approach', *Biometrics* **44**, 837–845.

Efron, Bradley & Robert J. Tibshirani (1994), *An Introduction to the Bootstrap*, Chapmann&Hall/CRC, London.

Engelmann, Bernd, Evelyn Hayden & Dirk Tasche (2003), 'Testing rating accuracy', *Risk* (1), 82–86.

Giesecke, Kay (2001), Default and information. *Jornal of Economic Dynamics and Control*, forthcoming.

Giesecke, Kay & Lisa Goldberg (2004), 'Forecasting default in the face of uncertainty', *Journal of Derivatives* **12**(1), 1–15.

Lehmann, Erich L. (1998), *Nonparametrics: Statistical Methods Based on Ranks*, Prentice Hall, London.

Poduri, Vijay (2004), 'Introduction to the Barra Default Probability model', *Horizon* (4), 2–10.