A Note on the Berkowitz Test with Discrete Distributions

Alfred Hamerle
Department of Statistics, University of Regensburg, Universitaetsstr. 31,
93053 Regensburg, Germany,
phone: +49 941 943 2588, e-mail: alfred.hamerle@wiwi.uni-regensburg.de

Kilian Plank
(corresponding author)
Department of Statistics, University of Regensburg, Universitaetsstr. 31,
93053 Regensburg, Germany,
phone: +49 941 943 2751, e-mail: kilian.plank@wiwi.uni-regensburg.de

A Note on the Berkowitz Test with Discrete Distributions

Abstract

Berkowitz (2001) suggested a powerful and popular density test based on a probability integral transformation (PIT). For the PIT to work properly the original distribution needs to be continuous. In this article, we show what problems can arise when the procedure is applied to discrete distributions. We suggest a simple modification so that the basic assumptions of the Berkowitz test are recovered.

KEYWORDS: *Credit Risk, Validation, Probability Integral Transform, Berkowitz Test*

1 Introduction

It is common practice to base risk assessments of mark-to-market portfolios on profit-/loss quantiles like Value at Risk (VaR) and to validate them by means of a binomial test (see e.g. Kupiec 1995, Crnkovic & Drachman 1996, Pritsker 1997, Lopez 1999). This test class, however, has the drawback that it disregards most of the distributional information and requires large sample sizes in order to be powerful. As large samples are often not available, Berkowitz (2001) suggests to use the more efficient class of density-based tests.

Contrary to the literature on VaR model validation there are only few corresponding contributions for credit portfolio models (Lopez & Saidenberg (2000) and Frerichs & Loeffler (2003)). The reasons for this research gap are lack of data and scarcity of default events which makes application of market risk validation methods extremely difficult. Hence, Lopez & Saidenberg (2000) suggested resampling techniques to mitigate the data problems. Later on, Frerichs & Loeffler (2003) applied the test procedure of Berkowitz (2001) to credit default/loss data. The authors report remarkable power results even for short time series. However, as we shall show in this article, the basic Berkowitz approach must be applied with caution in some situations. More precisely, the test is based on two probability transformations. The first of these transformations implicates a standard uniform distribution if the original distribution (the distribution of losses or defaults) is continuous. Otherwise, however, the result is not standard uniform and we show that this translates into biased test results. To resolve this problem we suggest a simple modification of the discrete probability transformation and show that this admits the usual χ^2 approximation.

The article is structured as follows. In the next section we shortly introduce the Berkowitz test approach. After that, we show the dangers of using a simple discrete probability transform. In Section 4 we suggest a modified probability transform and show how it improves the results. Section 5 concludes.

2 Berkowitz Test Procedure

The approach of Berkowitz (2001) involves testing whether hypothesized and true probability distribution coincide, i.e., $H_0 : F_{L_0}(l) = F_L(l)$. Here, F_L denotes the true cumulative distribution function (cdf) and F_{L_0} is the hypothesized cdf.

In a first step the observed loss time series $l_1, \ldots, l_T$ is transformed by its hypothetical probability integral $(F_{L_0}(l_t))_{t=1}^{T}$. Under H_0 the transformed observations are iid standard uniform, i.e.,

$$u_t = F_{L_0}(l_t) \sim \mathcal{U}(0,1), \tag{1}$$

where $\mathcal{U}(0,1)$ denotes the uniform distribution with support $[0,1]$.

In a second step the data $(u_t)_{t=1}^T$ are transformed once again via the inverse of the standard normal cdf. This admits application of a large set of (standard) normality tests. Finally, under H_0 we have that

$$z_t = \Phi^{-1}(u_t) \sim \mathcal{N}(0,1). \tag{2}$$

To summarize, if the first transformation is based on the true credit risk model, we get $u_t \sim \mathcal{U}(0,1)$ and $z_t \sim \mathcal{N}(0,1)$. Thus, we hypothesize that $H_0 : z_t \sim \mathcal{N}(0,1)$ for which Berkowitz (2001) proposes the classical likelihood ratio (LR) test. In case of rejection, there is an $\alpha \cdot 100$ percent probability (the test size) that the null hypothesis is true even so.

3 Discrete Probability Integral Transformation (PIT)

If the original distribution is discrete, u_t can of course not be standard uniform. To illustrate this shortly, take, for example, a credit portfolio default model represented by the random variable L counting the number of defaults over a certain period of time. After the first transformation we have $U = F_L(L)$. Since $L \geq 0$ and the number of defaults in a credit portfolio is usually low, a pronounced mass peak at $F_L(0)$ does exist, i.e., we have a singularity $\mathbb{P}(U = F_L(0)) > 0$. On the other hand, the interval $[0, F_L(0))$ has zero density, i.e., $\mathbb{P}(U < F_L(0)) = 0$.

More generally, let L denote a discrete random variable with K mass points. Let possible realisations of L be denoted by l so that the complete set of realisations in ascending order is

$$l^{(1)}, \ldots, l^{(K)}$$

Now we have that

$$\mathbb{P}\Big(L = l^{(k)}\Big) = \mathbb{P}\Big(U = F_L\left(l^{(k)}\right)\Big) > 0$$

but

$$\mathbb{P}\Big(l^{(k-1)} < L < l^{(k)}\Big) = \mathbb{P}\Big(F_L\left(l^{(k-1)}\right) < U < F_L\left(l^{(k)}\right)\Big) = 0$$

So the transfom of a discrete random variable is not standard uniform.

What does that imply for the Berkowitz test. Ties, i.e., multiple occurences of the same l_t and thus u_t and z_t, are particularly likely when the simulated portfolio is small or when default correlation is high. In the latter case, we have a massive peak at $L = 0$. But if the variance of L is high and if the portfolio is large the likelihood of several identical observations l_t is very low. In any case, if $U = F_L(L)$ is not standard uniform, $Z = \Phi^{-1}(F_L(L))$ is not standard normal and the $\chi^2(2)$ distribution is only a poor approximation for the distribution of the LR statistic. This problem cannot be resolved by increasing the sample size T. Indeed, application of the Berkowitz test on large samples, where the asymptotic χ^2 distribution is usually a good approximation, may be misleading when based on discrete probability integral transforms.

To illustrate this, we perform a simulation study. We use the standard Gaussian single-risk factor model as our credit portfolio model (e.g. Gordy 2000). The number of obligors is denoted by N. All obligors have the same default probability λ. Likewise, we assume homogeneous asset correlations ρ for all pairs of obligors. We consider default time series of length T.

The tested levels of N, λ, ρ, and T are shown in Table 1.

					n
T	5	10	100	500	4
N	100	1,000	10,000		3
ρ	5%	10%	20%		3
λ	0.1%	1%	5%		3

Table 1: Set of tested parameter levels.

Now we proceed as follows. We select a combination of parameter levels (T, N, λ, ρ) from Table 1 and simulate a default sample of length T. We draw $10,000$ such samples and calculate the LR statistic of the Berkowitz test and record the fraction of simulation runs where LR falls in the $\alpha = 10\%$ rejection area of a $\chi^2(2)$ distribution. When the discrete PIT has the hypothesized impact, we expect clear deviations from the 10% level.

The results are given in Table 2.

	$T = 5$			$T = 10$			$T = 100$			$T = 500$		
	$\rho = 5\%$	$\rho = 10\%$	$\rho = 20\%$	$\rho = 5\%$	$\rho = 10\%$	$\rho = 20\%$	$\rho = 5\%$	$\rho = 10\%$	$\rho = 20\%$	$\rho = 5\%$	$\rho = 10\%$	$\rho = 20\%$
N = 100												
$\lambda = 0.1\%$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000	1.0000	1.0000
$\lambda = 1\%$	0.4920	0.5115	0.7390	0.7325	0.8565	0.9855	1.0000	1.0000	1.0000	1.000	1.0000	1.0000
$\lambda = 5\%$	0.1885	0.1660	0.1875	0.1510	0.1520	0.1880	0.5525	0.6090	0.9195	0.997	1.0000	1.0000
N = 1,000												
$\lambda = 0.1\%$	0.4745	0.7300	0.9405	0.8110	0.9670	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\lambda = 1\%$	0.1595	0.1590	0.1775	0.1345	0.1490	0.1770	0.2495	0.3520	0.8595	0.7635	0.9525	1.0000
$\lambda = 5\%$	0.1590	0.1565	0.1455	0.1195	0.1280	0.1310	0.1095	0.1140	0.1055	0.1400	0.1410	0.2200
N = 10,000												
$\lambda = 0.1\%$	0.1695	0.1620	0.2235	0.1505	0.1465	0.3030	0.3070	0.5565	1.0000	0.8700	0.9985	1.0000
$\lambda = 1\%$	0.1420	0.1615	0.1630	0.1180	0.1120	0.1180	0.0930	0.0950	0.1195	0.1120	0.1160	0.2185
$\lambda = 5\%$	0.1575	0.1480	0.1560	0.1215	0.1140	0.1255	0.1060	0.1065	0.1040	0.1005	0.1110	0.0980

Table 2: Discrete PIT: Fraction of simulation runs (hitting rate) exceeding the $\chi^2(2)$ 90% quantile for varying levels of N, T, λ, and ρ.

First and foremost, we observe that deviations of the simulated hitting rate from the 10% ideal decrease top-down, i.e., from $N = 100$ to $N = 10,000$. The reason for this is that in a larger portfolio with not too low a PD the probability mass is distributed on more possible support points which reduces the probability of ties[1, 2]. The second observation from Table 2 has already been mentioned above. The hitting rate increases from left to right, i.e., larger samples deteriorate the applicability of the χ^2 distribution. Although counterintuitive at first sight the explanation is simply that in a larger sample the probability of ties is higher and ties lead to clear deviations from a Gaussian and χ^2, respectively[3]. Higher levels of ρ also implicate higher hitting rates. Again the reason is that a high level of correlation increases the probability of $L = 0$ and thus a large frequency of ties. Finally, higher levels of λ by and large decrease the hitting rate. This is due to the associated increase in the variance of the sample[4].

To sum up, we found that for small portfolios with low default rates and/or high correlation the basic Berkowitz test is not directly applicable. In the next section we suggest an approach to resolve this problem.

4 A Modified Probability Integral Transformation

As a solution to the aforementioned problems, we suggest to map u_t to a randomly drawn realisation from the left-adjacent interval.

Formally, let $u_t = F_L(l_t)$ and $u_t = u^{(k)}$, i.e., u_t is identical to the kth element in the ordered sequence of possible realisations $u^{(1)}, \ldots, u^{(K)}$ of $U = F_L(L)$. Then, we suggest to replace the pseudo observations $F_L(l_t)$ by random variables drawn from

$$U\left(u^{(k-1)}, u^{(k)}\right) \tag{3}$$

where $u^{(k-1)} = 0$ for $k = 1$.

It is obvious that this procedure establishes a uniform distribution over $(0, 1)$ since $\mathbb{P}\big(U = u^{(k)}\big) = u^{(k)} - u^{(k-1)}$ and our procedure warrants that this mass is uniformly distributed over the interval $\big(u^{(k-1)}, u^{(k)}\big)$.

The results of using this modified PIT on the hitting rates (i.e., again hits above the 90% $\chi^2(2)$ quantile) are given in Table 3.

Now we do no longer observe any differences in terms of the level of N, λ, and ρ. The reason is simply that the probability of ties is almost zero due to the modified PIT and the variance of the default distribution has no relevance. By contrast, we now observe the expected effect of larger samples, i.e., as T increases the hitting rates approach the asymptotic level

[1] In a large portfolio with a PD not too low the likelihood of two equal default counts (i.e., $\mathbb{P}(L_t = L_{t'}), t \neq t'$) is very low.

[2] Note that in Frerichs & Loeffler (2003), where the Berkowitz procedure is originally suggested for credit portfolio model validation, this problem is not mentioned. However, their simulation results are not influenced by discrete PIT bias as they use a large portfolio ($N = 10,000$).

[3] Ties lead to significant spikes in the distributions of U and Z and finally to a clear upward shift of the χ^2.

[4] Within common ranges the variance of portfolio loss increases as λ increases.

	$T = 5$			$T = 10$			$T = 100$			$T = 500$		
N = 100	$\rho = 5\%$	$\rho = 10\%$	$\rho = 20\%$	$\rho = 5\%$	$\rho = 10\%$	$\rho = 20\%$	$\rho = 5\%$	$\rho = 10\%$	$\rho = 20\%$	$\rho = 5\%$	$\rho = 10\%$	$\rho = 20\%$
$\lambda = 0.1\%$	0.1485	0.1470	0.1625	0.1180	0.1145	0.1400	0.0955	0.0980	0.1075	0.1010	0.0955	0.0945
$\lambda = 1\%$	0.1575	0.1520	0.1500	0.1260	0.1265	0.1310	0.1005	0.1075	0.0970	0.0975	0.1160	0.1100
$\lambda = 5\%$	0.1605	0.1410	0.1500	0.1170	0.1275	0.1155	0.1040	0.0925	0.0980	0.1035	0.1085	0.0985
N = 1,000												
$\lambda = 0.1\%$	0.1630	0.1475	0.1770	0.1230	0.1165	0.1330	0.1080	0.0875	0.1070	0.1100	0.1065	0.1060
$\lambda = 1\%$	0.1485	0.1555	0.1510	0.1200	0.1230	0.1185	0.0915	0.1005	0.1035	0.0850	0.1050	0.0920
$\lambda = 5\%$	0.1600	0.1560	0.1500	0.1190	0.1260	0.1325	0.0960	0.1090	0.0980	0.0995	0.1030	0.1080
N = 10,000												
$\lambda = 0.1\%$	0.1615	0.1470	0.1430	0.1325	0.1220	0.1220	0.1090	0.1055	0.0910	0.1075	0.1030	0.1015
$\lambda = 1\%$	0.1435	0.1625	0.1720	0.1180	0.1140	0.1300	0.0950	0.0960	0.0955	0.1040	0.0995	0.1030
$\lambda = 5\%$	0.1570	0.1475	0.1570	0.1220	0.1145	0.1270	0.1035	0.1075	0.1015	0.1005	0.1090	0.1055

Table 3: Modified Discrete PIT: Fraction of simulation runs exceeding the $\chi^2(2)$ 90% quantile for varying levels of N, T, λ, and ρ.

$\alpha = 10\%$. Even for very small samples (i.e., $T = 5$) the approximation error for the modified PIT is comparatively small. Compare, for example, the columns for $T = 100$ for the discrete and modified PIT. In the latter case, the rejection frequencies are between 9% and 11% while they range from 9% to 100% in the former case.

For a graphical comparison consider Figure 1 which compares a $\chi^2(2)$ distribution to the LR distribution based on discrete PIT and the LR distribution based on our modified PIT. It is obvious that while the modified PIT is close to the limiting distribution the simple discrete PIT is not. It is clearly shifted to the right. Comparing subfigure a) and b) one can clearly observe the aforementioned effect that larger portfolios reduce the error of using a discrete cdf.

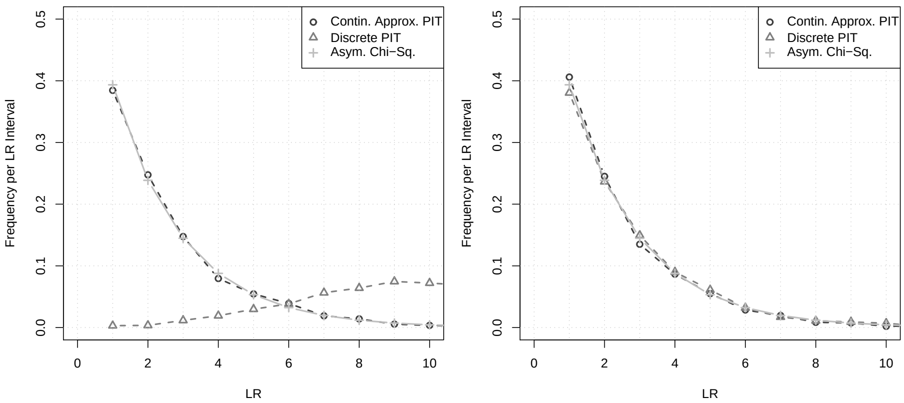
5 Conclusion

In this article we outline the dangers of using a discrete probability integral transformation within the context of the Berkowitz test. We show that larger portfolios, lower correlation, and higher PDs render the standard LR distributional approximation more reliable. By contrast, increasing the sample size may even deteriorate the error. We suggest a simple sampling technique to resolve this problem.

References

Berkowitz, J. (2001), 'Testing density forecasts, with applications to risk management', *Journal of Business and Economic Statistics* **19**(4), 465–474.

Crnkovic, C. & Drachman, J. (1996), 'Quality control', *Risk* (9), 139–143.



(a) $N = 1,000$ (b) $N = 10,000$

Figure 1: LR frequency polygons for modified PIT, discrete PIT and asymptotic $\chi^2(2)$ distribution. $T = 500$, $\lambda = 1\%$, $\rho = 10\%$ and (a) $N = 1,000$ and (b) $N = 10,000$.

Frerichs, H. & Loeffler, G. (2003), 'Evaluating credit risk models using loss density forecasts', *Journal of Risk* **5**(4), 1–23.

Gordy, M. B. (2000), 'A comparative anatomy of credit risk models', *Journal of Banking and Finance* **24**, 119–149.

Kupiec, P. (1995), 'Techniques for verifying the accuracy of risk measurement models', *Journal of Derivatives* **3**, 73–84.

Lopez, J. (1999), Methods for evaluating value-at-risk estimates, Technical Report 2, FRBSF Economic Review.

Lopez, J. & Saidenberg, M. (2000), 'Evaluating credit risk models', *Journal of Banking and Finance* **24**, 151–165.

Pritsker, M. (1997), 'Evaluating value-at-risk methodologies: Accuracy versus computational time', *Journal of Financial Services Research* **12**, 201–242.