Multi-period credit default prediction with time-varying covariates.

Orth, Walter

University of Cologne

17 March 2011

DISCUSSION PAPERS IN STATISTICS
AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 3/11

Multi-period credit default prediction with time-varying covariates

by

Walter Orth

First draft: March 17, 2011
This draft: November 22, 2011

DISKUSSIONSBEITRÄGE ZUR
STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

Albertus-Magnus-Platz, D-50923 Köln, Deutschland

Multi-period credit default prediction with time-varying covariates

Walter Orth*

First draft: March 17, 2011
This draft: November 22, 2011

Abstract

In credit default prediction models, the need to deal with time-varying covariates often arises. For instance, in the context of corporate default prediction a typical approach is to estimate a hazard model by regressing the hazard rate on time-varying covariates like balance sheet or stock market variables. If the prediction horizon covers multiple periods, this leads to the problem that the future evolution of these covariates is unknown. Consequently, some authors have proposed a framework that augments the prediction problem by covariate forecasting models. In this paper, we present simple alternatives for multi-period prediction that avoid the burden to specify and estimate a model for the covariate processes. In an application to North American public firms, we show that the proposed models deliver high out-of-sample predictive accuracy.

JEL classification: C41, C53, C58, G17, G32, G33

Keywords: Credit default, multi-period predictions, hazard models, panel data, out-of-sample tests

*Department of Statistics and Econometrics, University of Cologne; orth@statistik.uni-koeln.de
I would like to thank Karl Mosler and participants of the International Risk Management Conference 2011, Amsterdam, Credit Scoring and Credit Control XII, Edinburgh, and of the 2011 Conference of the German Statistical Society, Leipzig, for their comments.

1. Introduction

The need for multi-period credit default prediction arises naturally through the multi-period risk faced by creditors who lend with a maturity of multiple periods. Accordingly, rating agencies like Standard & Poor's assign their ratings – which can be viewed as ordinal default predictions – with a time horizon of up to five years (Standard & Poor's, 2010). Further, the Basel Committee on Banking Supervision states that "banks are expected to use a longer time horizon [than one year] in assigning ratings" (BCBS (Basel Committee on Banking Supervision), 2006, § 414). The importance of multi-period default predictions may further rise due to plans of the International Accounting Standards Board and the Basel Committee on Banking Supervision to base loss provisions upon the expected loss over the whole life of the credit portfolio (BCBS (Basel Committee on Banking Supervision), 2009). This would necessitate the estimation of multi-year default probabilities. However, the considerably larger fraction of studies in the huge default prediction literature only deals with a fixed and often short-term prediction horizon. For example, a standard approach in the literature is to estimate a discrete-time hazard model (which is basically equivalent to a binary panel model) with yearly data directly yielding one-year default probabilities (Shumway, 2001; Chava & Jarrow, 2004; Hillegeist et al., 2004; Männasoo & Mayes, 2009). In such a setting, default predictions for time horizons of more than one year are not directly available since the covariates are typically time-varying and their future evolution is unknown. Further, these models do not use all information if data are available on a quarterly or monthly basis. Therefore, even for one-year horizons, multi-period models are useful since they allow to use more information.

To overcome the problem of unknown future covariates, some authors have proposed a framework that augments the prediction problem by a covariate forecasting model. Duffie et al. (2007) first estimate a proportional hazard model with time-varying covariates and then forecast their vector of covariates by Gaussian panel vector autoregressions. Since a point forecast for the covariate vector is not sufficient due to the nonlinear nature of hazard models, the corresponding multivariate density forecast is then used to calculate the default probabilities by numerical methods. Similarly, Hamerle et al. (2007) estimate a discrete-time hazard model and subsequently develop a covariate forecasting model. They offer an interesting option to reduce the dimension of the problem by forecasting what they call the credit score, i.e. the inner product of the estimated parameter vector and the covariate vector, instead of the covariates themselves.[1] To compute the default probabilites, a Monte Carlo simulation of the paths of the credit scores is utilized. Approaches that involve covariate forecasting models have some drawbacks that make it worthwile to look for alternatives. First, there is a considerable additional burden in both model building and computing time. Second, there are purely statistical disadvantages since the econometrician is left with the choice between a large model containing many parameters – which may have low out-of-sample predictive power – and quite restrictive assumptions to reduce dimensionality. As examples for the latter note that the model of Duffie et al. (2007) contains only four covariates and that in the approach of Hamerle et al. (2007) equal credit scores lead to equal credit score forecasts regardless of the composition of the covariate vector. An advantage of the covariate forecasting approach – as compared to the models we will propose – is that it is capable of the analysis of *portfolio* credit risk since the covariate processes provide a model for the dependence of defaults. In contrast, we focus on *single-obligor* credit risk and present models designed for the purposes of rating and probability of default estimation hereby accentuating a multi-period prediction horizon. Ratings and default probabilities are – among other risk management applications – important inputs for credit portfolio models which may have a multi-period horizon as well.[2]

In this paper, we propose relatively simple models that deliver multi-period default predictions without a covariate forecasting model. We present the framework for these models in section 2 and discuss their estimation in section 3. In the empirical analysis of section 4, we apply these models to a dataset of North American public firms and show that they deliver high out-of-sample predictive accuracy. Section 5 concludes.

2. The econometric models

As a background to the subsequently presented approach we briefly address the study of Campbell et al. (2008). In their work, the authors estimate discrete-time hazard models lagging their time-varying covariates by s months, $s = 6, 12, 24, 36$.

[1]Besides the firm-specific covariates which are summarized in the credit score, Hamerle et al. (2007) also consider macroeconomic covariates which are summarized in a "macro score" and are similarly modeled.

[2]For instance, the popular industry model CreditRisk+ has a hold-to-maturity option and can thus be used as a multi-period model. For further discussion of multi-period credit portfolio models see Ebnöther & Vanini (2007); Hamerle et al. (2007); Rösch & Scheule (2007).

The authors point out that this approach can be extended by letting s run from 1 to H (H denoting the prediction horizon) meaning a stepwise increase of the lag index in the hazard regressions. Then, multi-period default probabilities can be calculated in closed form since the hazard rate in period $t+s$ is directly given as a function of the covariates in period t. Once the hazard rates for period $t+s$, $s=1,\dots,H$, are estimated, they are easily combined to give an estimate for the probability of a default event in the next H periods. However, estimation gets a bit cumbersome since one has to estimate H different parameter vectors which also increases the numbers of parameters substantially and thereby raises questions about out-of-sample predictive power. While Campbell et al. (2008) do not perform and validate such an extended approach, it nevertheless provides an interesting means to overcome the burden to specify a covariate forecasting model. Therefore, we will consider this approach in the empirical analysis. Before we do so, we will now introduce a framework that does not need a covariate forecasting model as well and thereby involves the estimation of just one parameter vector.

Let us first introduce the basic notation. We observe obligor i, $i=1,\dots,N$, for T_i periods thereby recording his default history and a vector of time-varying covariates x_{it}. Thus, we consider datasets that have a panel structure. Now, for each period t, $t=1,\dots,T_i$, define Y_{it} to be the lifetime (the time until default) of obligor i starting in period t. It follows that we observe T_i-1 partially overlapping lifetimes for each obligor since we have no information about the lifetime starting in the last period. In real datasets we will not observe the end of every lifetime, so that we have to define additionally the corresponding censoring indicator variable C_{it} which is zero in the case of no censoring, i.e. the lifetime ends with a default event, and one for censored lifetimes. We will specify our models in terms of the continuous-time hazard rate which is defined as

$$\lambda(y) = \lim_{\Delta y \to 0} \frac{P(y \leq Y < y + \Delta y \,|\, Y \geq y)}{\Delta y}\,. \tag{1}$$

The hazard rate measures the instantaneous risk of default. We choose the continuous-time specification since it is more common in the survival analysis literature and gives us a greater variety of models to choose from. Additionally, software packages usually offer more implementations for continuous-time hazard models.[3]

The idea behind the models we propose is as follows. Suppose that at a point in time t we want to predict the default probabilities for the next H periods using

[3] Discrete-time hazard models are often estimated by routines for binary panel models. However, this is not so easily done for our kind of models since we deal with a panel of lifetimes and not with a panel of binary variables.

the information, i.e. the covariates, we have at t. A simple solution is to specify the (time-t conditional) hazard rate in period $t+s$, $\lambda(t+s)$, as a function of the covariates in period t, x_{it}, and the "forecast time" s. For instance, if we choose a proportional hazard (PH) specification we would get

$$\lambda(t+s, x_{it}) = \lambda_0(s)\exp(\beta' x_{it})\,. \tag{2}$$

$\lambda_0(s)$ is called the baseline hazard and captures here the variation in the influence of the covariates over the forecast time. The variation of the hazard rate over s may also be interpreted as duration dependence. Note that the forecast time s is the analogon to the lag length in the approach of Campbell et al. (2008) outlined in the beginning of this section. There, the covariate effects are free to fluctuate for different s due to the repeated estimation of the model. In contrast, we impose a structure on the evolution of the covariate effects over the forecast time by integrating s as an argument into the functional form of the model. Importantly, the default probabilities are easily calculated in closed form:

$$P(Y_{it} \leq H) = 1 - \exp\left(-\int_0^H \lambda(t+s, x_{it})\,ds\right) \tag{3}$$

Note also the difference to the usual PH specification as, for instance, used in Duffie et al. (2007). There, the hazard rate in period $t+s$, $\lambda(t+s)$, is a function of the covariates in period $t+s$ leaving those models with the problem that the covariates are not known in $t+s$.

PH models have received great popularity not least because it is possible to estimate β without specifying the baseline hazard. This approach, developed by Cox (1972), can be followed by nonparametric estimation of the baseline hazard and is thus often called semiparametric. However, the PH model in our version implies that the hazard ratios for two obligors i and j, $\lambda(t+s, x_{it})/\lambda(t+s, x_{jt})$, are constant with respect to the forecast time s. There is evidence in the literature that this assumption is not realistic at least in the area of corporate credit (Fons, 1994). This can be easily seen by looking at tables of marginal default rates (the counterparts of hazard rates) for different rating grades.[4] There, it is quite obvious that the gap in the marginal default rates between firms with different ratings narrows with the prediction horizon. An intuitive interpretation is that the importance of the information in period t decays with the forecast time s. Fortunately, there is a class of hazard models that covers the case of converging hazard rates. Proportional odds

[4] Marginal default rates are easily derived from the cumulative default rates (the counterparts to default probabilities) regularly published by Standard & Poor's and Moody's.

(PO) models generally imply that the hazard ratios converge monotonically towards one (Bennett, 1983) where the convergence is with respect to the forecast time s in our setting. In PO models the survival odds and not the hazard rates of different obligors are constant multiples of each other over the whole forecast time. The most common PO specification is the log-logistic model where the conditional distribution of Y_{it} is assumed to be log-logistic. Then, in our framework the hazard rate is given by

$$\lambda(t+s, x_{it}) = \frac{\alpha \exp(\beta' x_{it})^{\alpha} s^{\alpha-1}}{1 + \{\exp(\beta' x_{it}) s\}^{\alpha}} . \tag{4}$$

Here, α determines the shape of the hazard curve. The cumulative distribution function evaluated at H (yielding the default probabilities) under this model is

$$P(Y_{it} \leq H) = 1 - \left[1 + \{\exp(\beta' x_{it}) H\}^{\alpha}\right]^{-1} . \tag{5}$$

While this model is fully parametric, there also exist semiparametric specifications for the PO model. For instance, Royston & Parmar (2002) use cubic splines for a flexible but smooth estimation of the baseline odds thereby achieving similar flexibility as in the Cox model. In our empirical analysis, we also experimented with this approach. However, this did not lead to improved predictive accuracy.[5] Thus, we do not document it further here.

3. Estimation

For the models we propose, the lifetimes starting at t, Y_{it}, are simply connected to the covariates in period t, x_{it}. Clearly, the multiple lifetimes of an individual obligor are not conditionally independent, i.e. Y_{it} is not conditionally independent from Y_{it^*}, $t \neq t^*$. Too see this, note that for instance Y_{it} already covers the lifetime $Y_{i,t+1}$ plus one additional period so that we have a sample of partially overlapping lifetimes. The reason why $Y_{i,t+1}$ is included although it is completely covered by Y_{it} is that the covariates vary from period t to $t+1$ and provide additional information. For the purpose of point estimation, it is possible to ignore the dependencies due to our overlapping sample and still to consistently estimate the parameters. This is a result from multivariate survival analysis (Lawless, 2003, Ch. 11) where the asymptotics only require that the lifetimes of different obligors are conditionally independent and that the number of obligors (N) approaches infinity. No assumptions are made

[5]The reason for this is arguably the fact that we measure predictive accuracy in terms of an accurate risk ordering of the obligors which is usually invariant to changes in the shape of the baseline hazard. See section 4 for further discussion.

about the dependence structure within the lifetimes of an individual obligor. Our pseudo log likelihood function is

$$\log L = \sum_{i=1}^{N} \sum_{t=1}^{T_i - 1} (1 - C_{it}) \cdot \log\{\lambda(t + Y_{it}, x_{it})\} + \log\{S(t + Y_{it}, x_{it})\}\,, \qquad (6)$$

where $S(\cdot) \equiv 1 - F(\cdot)$ is the so-called survival function referring to the cumulative distribution function $F(\cdot)$. In many applications the assumption of conditional independence of the lifetimes of different obligors may be at best approximately true because of common shocks which affect all obligors over the forecast time and which are not reflected in the covariates at the start of the lifetime. However, our approach can be justified by one theoretical and two practical arguments. First, results on Maximum Likelihood estimation under multi-way clustering indicate that an additional clustering (dependence) within the time dimension does not lead to inconsistency of our estimator at least if the time dimension is large as well (Cameron et al., 2011).[6] Second, dummy variables for each period should capture common shocks to a large extent. We experimented with this option but found – similar to Campbell et al. (2008) – no important effects on our results. And third, the high out-of-sample predictive power of our models (the central objective of our analysis) which will be reported in the upcoming section provides further support for our approach.

For the estimation of the log-logistic model, we simply substitute the definitions of the hazard rate and the survival function as given in the preceding section into Equation (6). For the semiparametric Cox model, we have to adjust our notation a bit. Suppose that we have r distinct values of uncensored lifetimes in our (overlapping) sample, $Y_{(1)}, \ldots, Y_{(r)}$. Further, denote by $R(Y_{(j)})$ the set of observations with a lifetime of at least $Y_{(j)}$ (those "at risk" at $Y_{(j)}$) and denote by $d_{(j)}$ the number of defaults at $Y_{(j)}$. Formally, $d_{(j)} = \sum_{i=1}^{N} \sum_{t=1}^{T_i-1} 1_{[Y_{it}=Y_{(j)}, C_{it}=0]}$. Further, we build the sum of the covariate vectors of all observations that ended with a default at $Y_{(j)}$ and denote this by $z_{(j)}$: $z_{(j)} = \sum_{i=1}^{N} \sum_{t=1}^{T_i-1} 1_{[Y_{it}=Y_{(j)}, C_{it}=0]} x_{it}$. Then, the pseudo log partial likelihood under the Breslow approximation (Breslow, 1974) for tied lifetimes is

$$\log L^p = \sum_{j=1}^{r} \beta' z_{(j)} - d_{(j)} \log \left\{ \sum_{l \in R(Y_{(j)})} \exp(\beta' x_l) \right\}\,. \qquad (7)$$

After β has been estimated, the baseline hazard can be estimated nonparametrically (Lawless, 2003, Ch. 7).

[6] In our empirical analysis the sample length is 352 months.

While we can consistently estimate our models under the working independence assumption, the dependencies due to overlapping lifetimes are not ignorable for covariance matrix estimation. In fact, unadjusted standard errors would be much too low. Instead, if we view all the lifetimes of an individual obligor as one cluster, we can apply cluster-robust covariance matrix estimation. Let $\widehat{V}_H = (-\delta^2 ln\, L/\delta\beta\delta\beta'|_{\widehat{\beta}})^{-1}$ be the conventional covariance matrix estimator based on the Hessian of the log likelihood function. Further, denote by $s_i(\widehat{\beta})$ the contribution of obligor i to the score vector. Then, the cluster-robust covariance matrix estimator is

$$\widehat{V}(\widehat{\beta}) = \widehat{V}_H \left(\sum_{i=1}^{N} s_i(\widehat{\beta})' s_i(\widehat{\beta}) \right) \widehat{V}_H \,. \tag{8}$$

Again, the estimator is consistent for $N \to \infty$. Note that for the Cox model the score contributions are not defined in the usual way. However, cluster-robust covariance matrix estimation is still possible. Details can be found in Lin (1994).

The implementation of our models is easy. If for every observation of the panel dataset the lifetime Y_{it} and the corresponding censoring indicator C_{it} is calculated, standard survival analysis routines can be employed. An option for cluster-robust standard errors is also available in many software packages. There is a final point to note about the definition of the lifetimes Y_{it}. Given that we usually assume a limited prediction horizon, H, it may be sensible to conduct an artificial censoring of the lifetimes at H thereby omitting possibly irrelevant information about what happened after H. For instance, with H equal to 60 months, we would set a value of 60 to all lifetimes larger than 60 together with a change in the censoring indicator if the lifetime ended with a default event before. Empirical tests show that while the differences are rather small it is indeed preferable to conduct such an additional censoring.

4. Empirical analysis

To construct our dataset for the empirical analysis, we have merged three different datasets all of them referring to North American public firms. First, we collect monthly Standard & Poor's rating and default data from Compustat. Consequently, default is in our study defined to be a default rating (D or SD) from Standard & Poor's. We then merge the default histories with quarterly balance sheet data from Compustat and monthly stock market data from CRSP. The balance sheet variables are taken to be constant over the months between financial statements so that the final dataset has monthly time intervals. Since there are on average two months between the end of the corresponding fiscal period and the reporting date we lag the balance sheet variables by two months so that the values of the variables should have been indeed available in each month. Further, following common practice and relying on the results from Chava & Jarrow (2004) we exclude financial firms (SIC codes 6000-6799). Finally, to eliminate the effect of outliers, we winsorized all variables at the 5th and 95th percentile. The final dataset consists of 339 222 firm-months from 3575 firms in the period from December 1980 until March 2010. We observe 498 different default events, but note that our definition of Y_{it} leads to 18 914 partially overlapping lifetimes in our sample that end with a default event.

For the selection of our covariates, we used the experience from studies based on similar datasets (Campbell et al., 2008; Chava & Jarrow, 2004; Shumway, 2001) to choose candidate variables. We consider market-based and accounting-based variables since recent research indicates that a combination of both delivers the highest predictive accuracy (Campbell et al., 2008; Agarwal & Taffler, 2008). We do not include macroeconomic covariates although it is likely that these have a significant impact on the hazard rate as well. The reason for doing so is that macroeconomic covariates are known to be important for the absolute level of default risk but only of little importance for the assessment of the relative risk of firms (Carling et al., 2007; Jacobson et al., 2008; Hamerle et al., 2006). We will concentrate on the latter (see also the discussion later in this section) and thus disregard macroeconomic variables.[7] The final specification of our models was derived by a backward selection approach that entailed the sequential reduction of the model containing all candidate variables. As the main criteria in the model selection process we used the Wald statistics and the associated p values of the covariates since we have to be careful with likelihood ratio tests and information criteria in a pseudo likelihood setting. Further, we looked for possible non-monotone effects of the variables on the hazard rate by grouping each covariate into quartiles and including the corresponding dummy variables into our model. We find strongly non-monotone effects for growth of total assets. Both high and low (highly negative) growth rates are associated with higher default risk. Therefore, our final model contains a dummy variable which is one if annual growth of total assets is in the upper or lower quartile and zero otherwise. The other covariates are quite standard and are used in this way or very similarly in the aforementioned studies. Note that liquidity variables were also tested but found to be insignificant. Finally, we checked the correlation between our covariates. There is no correlation above 0.5 so that multicollinear-

[7] For a comprehensive analysis of the influence of macroeconomic variables on the default risk of North American firms see Figlewski et al. (2012).

Table 1: Summary statistics for covariates

Name	Description	Mean	St.dev.	Min	Max
NITA	Net income over previous year / Total assets	.007	.020	-.155	.079
TLTA	Total liabilities / Total assets	.636	.168	.115	1
GRO	Dummy for extreme growth of total assets	.5	.5	0	1
RET	Excess one-year log stock return over S&P 500	-.029	.367	-1.317	1.220
VOLA	St. dev. of monthly log returns in previous year	.110	.063	.039	.298
SIZE	Log(market value / S&P 500 total market value)	-8.99	1.72	-13.27	-6.34

Table 2: Results from hazard regressions

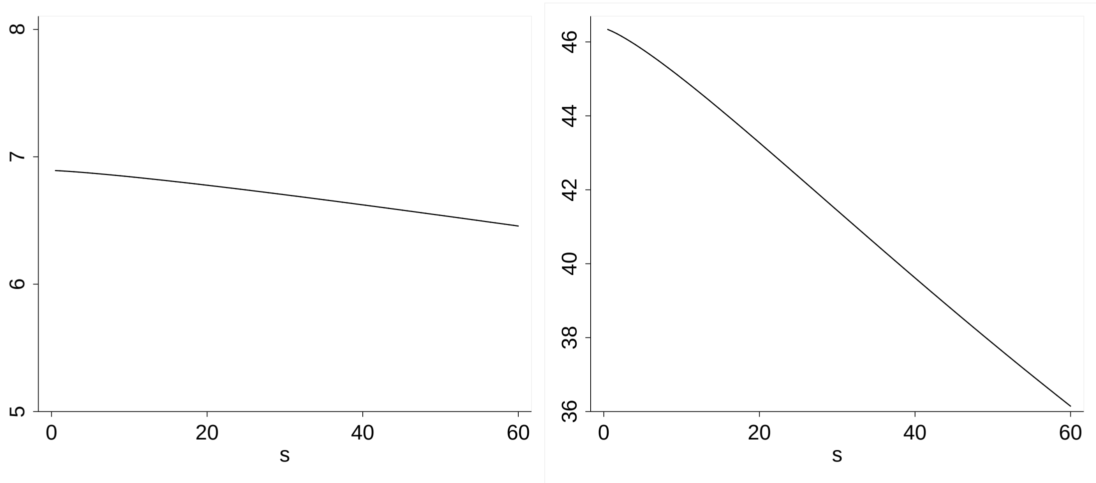
	Cox model (PH)		Log-logistic model (PO)	
	Coef.	Std. error	Coef.	Std. error
NITA	-5.598	(1.358)	-6.804	(1.271)
TLTA	2.426	(0.296)	2.311	(0.254)
GRO	0.212	(0.054)	0.184	(0.053)
RET	-0.826	(0.056)	-0.813	(0.053)
VOLA	6.142	(0.526)	6.062	(0.461)
SIZE	-0.374	(0.031)	-0.336	(0.027)
const.			11.992	(0.278)
α			1.255	(0.023)
firm-months	339 222		339 222	
Wald χ^2	2351.88		2415.62	

ity should not pose a problem. A brief description and summary statistics for all selected covariates are presented in Table 1.

We now turn to our estimation results. Table 2 shows the parameter estimates for the Cox model and the log-logistic model. The results refer to lifetimes which have been (additionally) censored at 60 months as described at the end of the previous section. All coefficients have the expected sign and are highly significant. As we estimate the Cox model with the partial likelihood approach, we do not estimate parameters for the baseline hazard (including the intercept) here. The results from the Cox model and the log-logistic model turn out to be quite similar. The goodness-of-fit is larger for the log-logistic model as can be seen by the corresponding Wald χ^2 statistics for the standard test that all coefficients of the covariates are zero.

Our primary motivation to estimate the log-logistic model was its property of de-

Figure 1: Evolution of hazard ratios



Hazard ratios of the upper and lower quartile firm (left) and upper and lower decile firm (right) derived from the log-logistic model

clining hazard ratios. To study the evolution of hazard rates over the forecast time we plotted in Figure 1 the hazard ratios of the upper and lower quartile firm on the left hand side and the hazard ratios of the upper and lower decile firm on the right hand side. For this calculation, we sorted all firms-months according to their risk (as measured by $x_{it}'\beta$) so that for the upper quartile firm, for instance, 25% of all firms-months are estimated to be less risky. The decrease of the hazard ratios is evident but seems to happen at a moderate pace. This does not surprise as we do not expect that the hazard rates of high-risk and low-risk firms approach each other very quickly. By comparing both curves we further see that more extreme hazard ratios decline more quickly. In the Cox model, the hazard ratios for the same quantiles are constant at 5.30 and 24.28, respectively.

We now proceed with the evaluation of the predictive power of our models. Following common practice we will focus on what is usually referred to as discriminative power, i.e. the ability of our models to provide an accurate rank order of the firms according to their risk. We will not evaluate the calibration of our models which concerns the levels of the predicted probabilities. The reason for this is that a recalibration of models is both possible and commonplace. In the credit risk area, such a recalibration is often done by classifying the obligors into rating classes and estimating default probabilities for these classes. In contrast, it is not possible to adjust the discriminative power of a model afterwards. To measure predictive accuracy, we will use the Accuracy Ratio and Harrell's C, a related measure from survival analysis which is advocated in Orth (2011) for credit risk applications.[8] The latter has a very similar interpretation as the Accuracy Ratio in that it is also bounded between -1 and 1 while the upper limit is reached by a hypothetical perfect model. The main differences are that Harrell's C can use censored observations and the timing of default events. We use our measures in the following way. For a given sample calendar month $\widetilde{t}$,[9] we calculate the Accuracy Ratio and Harrell's C for the predictions made in period $\widetilde{t}$ (and the corresponding lifetimes starting at $\widetilde{t}$). We do this in monthly steps for a range of values for $\widetilde{t}$ and then take a weighted average of our indices with the number of firms observed in each period as weights. We measure both in-sample and, more importantly, out-of-sample predictive power. In the in-sample part, $\widetilde{t}$ is ranging from December 1985 to March 2005 which covers all periods where the indices can be calculated. In the out-of-sample part, $\widetilde{t}$ is ranging from December 1995 to March 2005. There, each month the models are re-estimated using only the information available until period $\widetilde{t}$, a procedure known as a recursive estimation scheme. Stein (2004) calls it alternatively a walk-forward approach and argues that it is closest to the practical use of default prediction models.

Besides the Cox model and the log-logistic model we consider as competitors the stepwise lagging procedure (SLP) as outlined in the beginning of section 2 (using a logit specification for the discrete-time hazard rate as in Campbell et al., 2008) and Standard & Poor's Long Term Issuer Credit Ratings. As prediction horizons we choose one, three and five years. The results are shown in Table 3. We observe high predictive accuracy for all our models. While comparisons with other studies have to be taken with care note that Duffie et al. (2007) report out-of-sample Accuracy Ratios of 87% (one year) and 70% (five years) using a similar dataset thereby achieving less accuracy than our models.[10] Comparing our different specifications, we see that the log-logistic model performs best in every category. The Cox model is second-best in the out-of-sample part and similar to the SLP procedure in-sample. This difference is most likely due to the fact that the SLP approach is more highly parameterized and thus suffers more from out-of-sample instability than the other models. Standard & Poor's ratings throughout have the lowest predictive power

[8] Both measures are special cases of Somers' D which is again very closely related to Goodman and Kruskal's γ (Somers, 1962). The Accuracy Ratio is statistically equivalent to the Area under the ROC curve and also sometimes referred to as the Gini coefficient. See BCBS (Basel Committee on Banking Supervision) (2005) for an overview.

[9] $\widetilde{t}$ denotes calendar time whereas t as defined in section 2 is a firm-specific time index starting with 1 for each firm regardless of its entry date.

[10] Duffie et al. (2007) use a covariate forecasting approach as described in section 1 and state that their model is an improvement over available alternatives.

Table 3: Model performance statistics

Panel A: In-sample predictive accuracy

	Harrell's C			Accuracy Ratio		
Prediction horizon (months)	12	36	60	12	36	60
log-logistic	.9011	.8071	.7593	.9086	.8283	.7931
Cox	.9003	.8061	.7580	.9077	.8274	.7917
SLP	.9004	.8065	.7571	.9078	.8279	.7910
S&P	.8264	.7616	.7284	.8353	.7929	.7784

Panel B: Out-of-sample predictive accuracy

	Harrell's C			Accuracy Ratio		
Prediction horizon (months)	12	36	60	12	36	60
log-logistic	.8862	.7672	.7104	.8939	.7864	.7436
Cox	.8840	.7628	.7059	.8917	.7819	.7389
SLP	.8829	.7586	.6993	.8906	.7785	.7338
S&P	.8149	.7338	.6943	.8234	.7625	.7417

with the exception of the out-of-sample five-year Accuracy Ratio. The gains from our models as compared to S&P are highest for the shorter horizons. This is similar to findings in related studies and is also in line with the common perception that rating agencies are not making the most efficient use of short-term relevant information.

We now go on to analyze if the differences in out-of-sample predictive power between our competing predictors are statistically significant. We choose the bootstrap as a robust tool for this purpose. Due to the dependencies in our data we have to choose a bootstrap procedure that accounts for this. To be specific, we need a bootstrap procedure which is based on resampling with replacement from approximately independent units. This can be achieved by resampling from the set of firms instead of the set of firm-months.[11] Then, the standard bootstrap formulas can be applied. Such an approach is actually a version of the cluster bootstrap (Field & Welsh, 2007) if we view again all observations of one obligor as one cluster. By resampling from our out-of-sample predictors and the associated lifetimes we can perform bootstrap hypothesis tests for the null that two models have the same predictive power.

The results of Table 4 show that the log-logistic model is a significant improvement

[11] In a related application, Hanson & Schuermann (2006) employ this procedure as well.

Table 4: Bootstrap hypothesis tests for out-of-sample predictive accuracy

Prediction horizon of 12 months

	Harrell's C				Accuracy Ratio			
	log-l.	Cox	SLP	S&P	log-l.	Cox	SLP	S&P
log-l.	.	.002	.002	.001	.	.001	.001	.001
Cox		.	.009	.001		.	.008	.001
SLP			.	.001			.	.001
S&P				.				.

Prediction horizon of 36 months

	Harrell's C				Accuracy Ratio			
	log-l.	Cox	SLP	S&P	log-l.	Cox	SLP	S&P
log-l.	.	.001	.001	.012	.	.001	.001	.068
Cox		.	.009	.022		.	.029	.135
SLP			.	.056			.	.217
S&P				.				.

Prediction horizon of 60 months

	Harrell's C				Accuracy Ratio			
	log-l.	Cox	SLP	S&P	log-l.	Cox	SLP	S&P
log-l.	.	.001	.001	.223	.	.001	.001	.871
Cox		.	.001	.383		.	.009	.816
SLP			.	.700			.	.575
S&P				.				.

The table contains p values for the null hypothesis that the population values of the indices (i.e., the Accuracy Ratio or Harrell's C) for two predictors are equal which is tested against the two-sided alternative. The test refers to the results of Table 3, Panel B. The number of bootstrap replications is $B = 999$. If ΔI denotes the difference of two indices and * refers to a bootstrap sample, the formula for the p values is $p = \{1 + \#(|\Delta I^* - \Delta I| \geq |\Delta I|)\}/\{1 + B\}$. See Davison & Hinkley (1997) for details.

over all alternatives with the exception of S&P ratings at the five-year time horizon. Further, the stepwise lagging procedure performs significantly worse than the more parsimonious Cox and log-logistic models. This result holds regardless of the prediction horizon and the accuracy measure used. Our findings give rise to the following two main interpretations. On the one hand, we see that it pays off to choose a parsimonious model with relatively few parameters. This is of course a common finding especially in the forecasting literature. On the other hand, we observe that it is worthwhile to thoroughly analyze the structure imposed by the functional form. Here, the more realistic assumption of converging hazard rates of the log-logistic model as opposed to the constant hazard ratio assumption of the Cox model leads to a significantly higher predictive accuracy.

5. Conclusions

In this paper we have derived and estimated simple but accurate models for multi-period credit default predictions. The relative simplicity stems mainly from the facts that in our approach no covariate forecasting model is needed and that the models can be estimated in one step. Our approach further has the advantage that extensions are quite straightforward. Such possibilities include semi- or nonparametric specifications and forecast combinations/model averaging. Moreover, extensive out-of-sample tests as we have done are computationally affordable. All these things would be a much greater challenge in more complex settings.

We apply our models to corporate defaults, but our approach is generally useful for predictive hazard models if the data have a panel structure and if the time intervals are shorter than the prediction horizon. Obvious related applications in the credit risk area include models of sovereign risk or models based on mortgage data with time-varying characteristics.

Finally, the sparseness of multi-period default prediction models in the literature partly seems to be based on the seemingly high effort needed for model building and computing as compared to single-period models. With relatively easy estimation methods at hand and due to their economic importance, multi-period models deserve to gain popularity.

References

Agarwal, V. & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, 32, 1541–1551.

BCBS (Basel Committee on Banking Supervision) (2005). *Studies on the Validation of Internal Rating Systems*. Basel Committee on Banking Supervision Working Paper No. 14.

BCBS (Basel Committee on Banking Supervision) (2006). *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*.

BCBS (Basel Committee on Banking Supervision) (2009). *Guiding Principles for the Replacement of IAS 39*. Available at http://www.bis.org/publ/bcbs161.htm.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2, 273–277.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics*, 30, 89–100.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business and Economic Statistics*, 29 (2), 238–249.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *Journal of Finance*, 63, 6, 2899–2939.

Carling, K., Jacobson, T., Linde, J., & Roszbach, K. (2007). Corporate credit risk modeling and the macroeconomy. *Journal of Banking & Finance*, 31, 845–868.

Chava, S. & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8, 537–569.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society Series B*, 34, 2, 187–220.

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge Unversity Press.

Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83, 635–665.

Ebnöther, S. & Vanini, P. (2007). Credit portfolios: What defines risk horizons and risk measurement? *Journal of Banking & Finance*, 31, 3663–3679.

Field, C. A. & Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society Series B*, 69, 369–390.

Figlewski, S., Frydman, H., & Liang, W. (2012). Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance*, 21(1), 87–105.

Fons, J. S. (1994). Using default rates to model the term structure of credit risk. *Financial Analysts Journal*, 50 (5), 25–32.

Hamerle, A., Jobst, R., Liebig, T., & Rösch, D. (2007). Multiyear risk of credit losses in SME portfolios. *Journal of Financial Forecasting*, 1(2), 1–29.

Hamerle, A., Liebig, T., & Scheule, H. (2006). Forecasting credit event frequency – empirical evidence for West German firms. *Journal of Risk*, 9 (1), 75–98.

Hanson, S. & Schuermann, T. (2006). Confidence intervals for probabilities of default. *Journal of Banking & Finance*, 30, 2281–2301.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9, 5–34.

Jacobson, T., Kindell, R., Linde, J., & Roszbach, K. (2008). *Firm Default and Aggregate Fluctuations*. Sveriges Riksbank Working Paper Series No. 226.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Hoboken, NJ: Wiley.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine*, 13, 2233–2247.

Männasoo, K. & Mayes, D. G. (2009). Explaining bank distress in Eastern European transition economies. *Journal of Banking & Finance*, 33 (2), 244–253.

Orth, W. (2011). The predictive accuracy of credit ratings: Measurement and statistical inference. *International Journal of Forecasting*. Forthcoming. doi:10.1016/j.ijforecast.2011.07.004.

Royston, P. & Parmar, M. K. B. (2002). Flexible parametric-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21, 2175–2197.

Rösch, D. & Scheule, H. (2007). Multi-year dynamics for forecasting economic and regulatory capital in banking. *Journal of Credit Risk*, 3 (4), 113–134.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*, 74, 101–124.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review*, 27(6), 799–811.

Standard & Poor's (2010). *The Time Dimension of Standard & Poor's Credit Ratings*. Global Credit Portal, September 22, 2010.

Stein, R. M. (2004). *Benchmarking Default Prediction Models*. Moody's KMV technical report No. 030124. Available at http://www.moodyskmv.com/research/.