Multi-period credit default prediction with time-varying covariates

Walter Orth
University of Cologne, Department of Statistics and Econometrics

Overview

Motivation

- Problem:
  Default prediction with a flexible **multi-period** time horizon
- Objective:
  Development of a model with high (out-of-sample) discriminatory power, i.e. a model that *ranks* the obligors according to their default probabilities accurately.

Multi-period vs. single-period default prediction models

- Only a small fraction of the default prediction literature deals with multi-period predictions.

Multi-period vs. single-period default prediction models

- Only a small fraction of the default prediction literature deals with multi-period predictions.
- Common approach: Modelling one-year default probabilities by estimating a discrete-time hazard model with covariates lagged by one year.

Multi-period vs. single-period default prediction models

- Only a small fraction of the default prediction literature deals with multi-period predictions.
- Common approach: Modelling one-year default probabilities by estimating a discrete-time hazard model with covariates lagged by one year.
- Such a model
  - cannot be easily extended to more than one year because the future values of the covariates are unknown.
  - does not use all information if data are quarterly/monthly.

Basic notation

- Y: Lifetime / Time until default
  Definition of hazard rate in discrete time:

$$\lambda(y) = P(Y = y | Y \geq y)$$

  Definition in continuous time:

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

Basic notation

- Y: Lifetime / Time until default
  Definition of hazard rate in discrete time:
  $$\lambda(y) = P(Y = y | Y \geq y)$$
  Definition in continuous time:
  $$\lambda(y) = \lim_{\Delta y \to 0} \frac{P(y \leq Y < y + \Delta y \,|\, Y \geq y)}{\Delta y}$$

- We observe obligor i, $i = 1, \dots, n$, for t_i periods recording the default history and time-varying covariates x_{it} ($\Rightarrow$ panel data).

Basic notation

- Y: Lifetime / Time until default
  Definition of hazard rate in discrete time:

$$\lambda(y) = P(Y = y | Y \geq y)$$

  Definition in continuous time:

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

- We observe obligor i, $i = 1, \ldots, n$, for t_i periods recording the default history and time-varying covariates x_{it} ($\Rightarrow$ panel data).
- Y_{it}: Lifetime of obligor i starting at t

Basic notation

- Y: Lifetime / Time until default
  Definition of hazard rate in discrete time:
  $$\lambda(y) = P(Y = y | Y \geq y)$$
  Definition in continuous time:
  $$\lambda(y) = \lim_{\Delta y \to 0} \frac{P(y \leq Y < y + \Delta y \,|\, Y \geq y)}{\Delta y}$$
- We observe obligor i, $i = 1, \ldots, n$, for t_i periods recording the default history and time-varying covariates x_{it} ($\Rightarrow$ panel data).
- Y_{it}: Lifetime of obligor i starting at t
- Main economic interest: Default probability $P(Y_{it} \leq H)$ for various prediction horizons H given the information available until t

Approaches that involve covariate forecasting

Continuous-time model of Duffie et al. (JFE 2007):

$$\lambda(t, x_{it}) = exp(\beta' x_{it})$$

The (four) covariates are modelled with Gaussian panel vector autoregressions. The probability of default until time H is given by

$$P(Y_{it} \leq H) = 1 - E\left[exp\left(-\int_0^H \lambda(t+s, X_{i,t+s})\, ds\right)\right] ,$$

which is approximated by numerical methods.

A similar approach that also involves the estimation of a covariate forecasting model is given by Hamerle et al. (JFF 2006).

Drawbacks of approaches with covariate forecasting

- Complexity: A multivariate density forecast for a vector of covariates over multiple periods is needed.

Drawbacks of approaches with covariate forecasting

- Complexity: A multivariate density forecast for a vector of covariates over multiple periods is needed.
- This complexity either results in highly parameterized models (that may perform poorly out of sample) or very restrictive assumptions in order to reduce dimensionality.

Drawbacks of approaches with covariate forecasting

- Complexity: A multivariate density forecast for a vector of covariates over multiple periods is needed.
- This complexity either results in highly parameterized models (that may perform poorly out of sample) or very restrictive assumptions in order to reduce dimensionality.
- Computational burden since closed-form solutions are usually not available.

Stepwise lagging of covariates

Campbell et al. (JF 2008) estimate discrete-time hazard models lagging the covariates by s months, $s = 6, 12, 24, 36$:

$$\lambda(t+s, x_{it}) = [1 + exp(\beta_s' x_{it})]^{-1}$$

If we extend this idea and apply a stepwise lagging procedure (SLP) by estimating the model for every s, $s = 1, \ldots, H$, the H-period default probabilities are given by:

$$P(Y_{it} \leq H) = 1 - \prod_{s=1}^{H} [1 - \lambda(t+s, x_{it})]$$

Overview

- Introduction
- Approaches in the literature
- **The proposed models**
- Empirical analysis
- Conclusions

We propose to specify the hazard rate in period $t + s$ as a function of the forecast time s and the covariates in period t. For instance, within the proportional hazard specification we get

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

We propose to specify the hazard rate in period $t + s$ as a function of the forecast time s and the covariates in period t. For instance, within the proportional hazard specification we get

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

- In this model, each covariate vector x_{it} in our panel is connected to the corresponding lifetime Y_{it}.

We propose to specify the hazard rate in period $t+s$ as a function of the forecast time s and the covariates in period t. For instance, within the proportional hazard specification we get

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

- In this model, each covariate vector x_{it} in our panel is connected to the corresponding lifetime Y_{it}.
- Note that conventional models would be specified as $\lambda(t, x_{it}) = \lambda_0(t) exp(\beta' x_{it})$ leaving those models with the problem that the covariates are not known in $t+s$.

We propose to specify the hazard rate in period $t + s$ as a function of the forecast time s and the covariates in period t. For instance, within the proportional hazard specification we get

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

- In this model, each covariate vector x_{it} in our panel is connected to the corresponding lifetime Y_{it}.
- Note that conventional models would be specified as $\lambda(t, x_{it}) = \lambda_0(t) exp(\beta' x_{it})$ leaving those models with the problem that the covariates are not known in $t + s$.
- The H-period default probabilities are easily calculated as $P(Y_{it} \leq H) = 1 - exp(-\int_0^H \lambda(t + s, x_{it}) ds)$.

We propose to specify the hazard rate in period $t+s$ as a function of the forecast time s and the covariates in period t. For instance, within the proportional hazard specification we get

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

- In this model, each covariate vector x_{it} in our panel is connected to the corresponding lifetime Y_{it}.
- Note that conventional models would be specified as $\lambda(t, x_{it}) = \lambda_0(t) exp(\beta' x_{it})$ leaving those models with the problem that the covariates are not known in $t+s$.
- The H-period default probabilities are easily calculated as $P(Y_{it} \leq H) = 1 - exp(-\int_0^H \lambda(t+s, x_{it}) ds)$.
- In our specification we only have to estimate the model **once** in contrast to the stepwise lagging approach.

Estimation

- Clearly, the lifetimes Y_{it} are not (conditionally) independent. For instance, Y_{it} already covers the lifetime $Y_{i,t+1}$ plus one additional period.

Estimation

- Clearly, the lifetimes Y_{it} are not (conditionally) independent. For instance, Y_{it} already covers the lifetime $Y_{i,t+1}$ plus one additional period.
- However, we can consistently ($n \to \infty$) estimate our model treating the observations as independent. Let C_{it} be the censoring indicator corresponding to Y_{it}.The pseudo log likelihood function is given by

$$log\ L = \sum_{i=1}^{n} \sum_{t=1}^{t_i - 1} (1 - C_{it}) \cdot log(\lambda(t + Y_{it}, x_{it})) + log(1 - F(t + Y_{it}, x_{it}))$$

Estimation

- Clearly, the lifetimes Y_{it} are not (conditionally) independent. For instance, Y_{it} already covers the lifetime $Y_{i,t+1}$ plus one additional period.
- However, we can consistently ($n \to \infty$) estimate our model treating the observations as independent. Let C_{it} be the censoring indicator corresponding to Y_{it}.The pseudo log likelihood function is given by

$$log\ L = \sum_{i=1}^{n} \sum_{t=1}^{t_i - 1} (1 - C_{it}) \cdot log(\lambda(t + Y_{it}, x_{it})) + log(1 - F(t + Y_{it}, x_{it}))$$

- For valid inference, we have to adjust the standard errors for the clustering within the observations of each obligor.

The log-logistic model

- The proportional hazards (PH) specification given above assumes that hazard ratios are constant over forecast time. However, several studies find that hazard rates of different firms tend to approach each other.

The log-logistic model

- The proportional hazards (PH) specification given above assumes that hazard ratios are constant over forecast time. However, several studies find that hazard rates of different firms tend to approach each other.
- In contrast, proportional odds (PO) models imply that the hazard ratios converge monotonically towards one (Bennett, AS 1983).

The log-logistic model

- The proportional hazards (PH) specification given above assumes that hazard ratios are constant over forecast time. However, several studies find that hazard rates of different firms tend to approach each other.
- In contrast, proportional odds (PO) models imply that the hazard ratios converge monotonically towards one (Bennett, AS 1983).
- The most common PO model is the log-logistic model where the hazard rate is given by

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

  The CDF evaluated at H (which gives the default probabilities) is

$$P(Y_{it} \leq H) = \frac{1}{1+[exp(\beta' x_{it})H]^{\alpha}}$$

Overview

The dataset

- Default histories, balance sheet and stock market variables for North American public firms from Compustat and CRSP

The dataset

- Default histories, balance sheet and stock market variables for North American public firms from Compustat and CRSP
- Excluding financial firms we have 339,222 non-missing firm-months and 3575 firms in the time from December 1980 until March 2010.

The dataset

- Default histories, balance sheet and stock market variables for North American public firms from Compustat and CRSP
- Excluding financial firms we have 339,222 non-missing firm-months and 3575 firms in the time from December 1980 until March 2010.
- We observe 498 different default events, but our definition of Y_{it} leads to 18,914 lifetimes in our sample that end with a default.

Selection of regressors

Using a general-to-specific variable selection approach based on candidate variables taken from related studies we end up with the following set of regressors:

- Profitability: Net Income / Total Assets (NITA)
- Leverage: Total Liabilities / Total Assets (TLTA)
- Growth: Dummy for very high or very low growth of Total Assets (GRO)
- Stock return: Excess one-year log return over S&P 500 (RET)
- Volatility: Standard deviation of monthly log returns over previous year (VOLA)
- Size: Log of market value relative to total market value of S&P 500 (SIZE)

Estimation results

	Cox model (PH)		Log-logistic model	
	Coef.	Std. Err.	Coef.	Std. Err.
NITA	-5.60	(1.36)	-6.80	(1.27)
TLTA	2.43	(0.30)	2.31	(0.25)
GRO	0.21	(0.05)	0.18	(0.05)
RET	-0.83	(0.06)	-0.81	(0.05)
VOLA	6.14	(0.53)	6.06	(0.46)
SIZE	-0.37	(0.03)	-0.34	(0.03)
const.			11.99	(0.28)
α			1.26	(0.02)

Evaluation of predictive power

- Competitors: Cox model, log-logistic model, stepwise lagging procedure (SLP) and Standard & Poor's Long Term Issuer Credit Ratings

Evaluation of predictive power

- Competitors: Cox model, log-logistic model, stepwise lagging procedure (SLP) and Standard & Poor's Long Term Issuer Credit Ratings
- Prediction horizons: 1, 3 and 5 years

Evaluation of predictive power

- Competitors: Cox model, log-logistic model, stepwise lagging procedure (SLP) and Standard & Poor's Long Term Issuer Credit Ratings
- Prediction horizons: 1, 3 and 5 years
- For a given sample month t, we calculate the Accuracy Ratio and Harrell's C for the out-of-sample predictions made at t. We then take a weighted average of the time series of indices using the number of firms observed in t as weights.

Evaluation of predictive power

- Competitors: Cox model, log-logistic model, stepwise lagging procedure (SLP) and Standard & Poor's Long Term Issuer Credit Ratings
- Prediction horizons: 1, 3 and 5 years
- For a given sample month t, we calculate the Accuracy Ratio and Harrell's C for the out-of-sample predictions made at t. We then take a weighted average of the time series of indices using the number of firms observed in t as weights.
- Range of t: December 1995 - March 2005.

Out-of-sample predictive power

	1 year		3 years		5 years	
	AR	C	AR	C	AR	C
log-logistic	.8939	.8862	.7864	.7672	.7436	.7104
Cox	.8917	.8840	.7819	.7628	.7389	.7059
SLP	.8906	.8829	.7785	.7586	.7338	.6993
S&P	.8234	.8149	.7625	.7338	.7417	.6943

Testing for significant differences

- Using the bootstrap we tested for significant differences in out-of-sample predictive accuracy. The tests yield the following main results:
  - The log-logistic model has significantly more predictive power ($\alpha = .1$) than all alternatives and at all horizons with the exception of Standard & Poor's for the 5-year horizon.

Testing for significant differences

- Using the bootstrap we tested for significant differences in out-of-sample predictive accuracy. The tests yield the following main results:
  - The log-logistic model has significantly more predictive power ($\alpha = .1$) than all alternatives and at all horizons with the exception of Standard & Poor's for the 5-year horizon.
  - The stepwise lagging procedure (SLP) is significantly worse ($\alpha = .05$) than both the log-logistic and the Cox model under all measures and horizons. This is probably due to overparameterization.

Main results

- We have derived a simple modeling approach for multi-period default predictions that does not involve the problem of forecasting covariates.

Main results

- We have derived a simple modeling approach for multi-period default predictions that does not involve the problem of forecasting covariates.
- The empirical part showed that our approach has high out-of-sample predictive power.

Main results

- We have derived a simple modeling approach for multi-period default predictions that does not involve the problem of forecasting covariates.
- The empirical part showed that our approach has high out-of-sample predictive power.
- The proportional odds model in the log-logistic specification was shown to fit significantly better in our application than the 'workhorse' of survival analysis, the Cox proportional hazards model.