The hidden Risk: Estimation of Unobservable Credit Risk

Statistical Learning Research Seminar (2010)

Stefan Kerbl

joint project work with Michael Sigmund

8th September 2010

What we do

We try to deepen our understanding of **how economic environment affects aggregate credit risk (ACR),** that is

(i) finding models which can identify factors that influence ACR and

(ii) using these models to forecast ACR,

according to the underlying assumption that ACR, expressed as aggregate default probability, is a function of the economic environment, $PD_t = f(EE_t)$.

Motivation

Why is estimation and forecasting of ACR of importance?
From *banks*' perspective it's highly relevant for

- banks' capital planning
- banks' stress testing programs
- rating models and their philosophy (*PIT* vs *TTC*)
- ...

From *systemic* perspective it's highly relevant for

- (ad hoc) assessment of the resilience of the financial sector
- system wide stress testing
- holistic macro economic modelling that incorporates credit risk
- ...

Orthodox approaches and challenges

Estimation and forecasting of ACR approached by orthodox methods rely on

- multivariate linear regression models
- vector autoregressive models
- factor models

However, with their application two aching issues have to be tolerated

1. A huge list of macro variables are available to explain credit risk. **Selecting regressors** becomes even more challenging when taking the possible lag structure into account. $\Rightarrow$ How to chose?
2. At the same time there is mounting evidence of ***unobserved*** **credit cycles**.

Tackling the first issue: employment of statistical learning methods

To effectively choose from the huge number of possible regressors we make use of the following techniques:

- Best-Subset Selection p. 57
- Shrinkage Methods
  - Ridge Regression p. 61
  - The Lasso p. 68
  - Elastic Net p. 73

(pages refer to [Hastie et al., 2009])

These methods are state of the art in model selection techniques.

Best-Subset Selection

... minimizes the in-sample prediction error, i.e. sum of sq. residuals, for models with $k \in \{0, 1, \ldots, p\}$ regressors.
Output is the optimal model for specified k.
in R:

```
package:leaps
command:regsubsets
methods:"exhaustive","backward","forward","seqrep"
```

(package *leaps* looks somewhat unfinished)

Shrinkage Methods

... minimizes *modified* in-sample prediction error including *parameter size penalty*.

$$\tilde{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2 + \lambda \sum_{j=1}^{p} |\beta_j|^q \right\}$$

(question to p. 63)

in R:

```
package: elasticnet (and lars)
command: enet
parameter: lambda
```

(lambda confusing: not lambda in [Hastie et al., 2009] but quadratic penalty: $\lambda\,\beta_j^2 + (1-\lambda)|\beta_j|$)

Tackling the second issue: Hidden Markov Models

Several recently published articles find unobserved components within credit risk as crucial: [Banachewicz et al., 2008], [Bruche and Gonzalez-Aguado, 2010] and [Koopman et al., 2008]. However, there is no precisely defined source. Surveying literature we find three promising ideas:

1. leverage cycle: leverage and/or solvency ratios of creditors, [Geanakoplos, 2010]
2. credit supply side driven effects: too lenient credit standards in phases of underestimated risks materializes late, capital buffer of banks
3. cyclical default correlation: contagion and spillover effects, [Giesecke, 2004]

Application: Kalman filter, smoother and parameter estimation

We estimate the unobserved component in the state spaced model,

$$x_t = \phi x_{t-1} + w_t \quad (1)$$
$$y_t = A x_t + \Gamma u_t + v_t, \quad (2)$$

where $w_t \sim N(0, Q)$, $v_t \sim N(0, R)$ and x_t is unobserved.
in R: While there are packages for Kalman filtering and smoothing (e.g. MARSS), we did not find estimation procedures for model parameters.
$\Rightarrow$ Own code: estimation of parameters according to ML following [Shumway and Stoffer, 2006] and [Durbin and Koopman, 2001]

Examples in R

Examples in R:

- data structure
- model selection
- unobserved component estimation

Open issue: How to effectively combine elastic net and unobserved component estimation?
Until now: elastic net only selects exogenous, but shrinkage is lost in re-estimation with unobserved.

Banachewicz, K., Lucas, A., and Van der Vaart, A. (2008).
Modelling portfolio defaults using hidden Markov models with covariates.
Econometrics Journal, 11(1):155–171.

Bruche, M. and Gonzalez-Aguado, C. (2010).
Recovery rates, default probabilities, and the credit cycle.
Journal of Banking and Finance, 34(4):754–764.

Durbin, J. and Koopman, S. (2001).
Time Series Analysis by State Space Methods, volume 24 of *Oxford Statistical Science Series*.
Oxford University Press.

Geanakoplos, J. (2010).
The Leverage Cycle.
In Acemoglu, D., Rogoff, K., and Woodford, M., editors, *NBER Macroeconomics Annual 2009*, volume 24 of

NBER Book Series NBER Macroeconomics Annual, chapter 1, pages 1–65. University of Chicago Press.

Giesecke, K. (2004).
Correlated default with incomplete information.
Journal of Banking & Finance, 28(7):1521–1545.

Hastie, T., Tibshirani, R., and Friedman, J. (2009).
The Elements of Statistical Learning: Data Mining, Inference, and Prediction.
Springer Series in Statistics. Springer, 2nd edition.

Koopman, S. J., Lucas, A., and Schwaab, B. (2008).
Forecasting cross-sections of frailty-correlated default.
Tinbergen Institute Discussion Papers 08-029/4, Tinbergen Institute.

Shumway, R. H. and Stoffer, D. S. (2006).
Time Series Analysis and Its Applications With R Examples.
Springer, New York.
ISBN 978-0-387-29317-2.