Correlation across latent variables in credit risk models: a direct inference from default rates

Fernando Moreira
University of Edinburgh

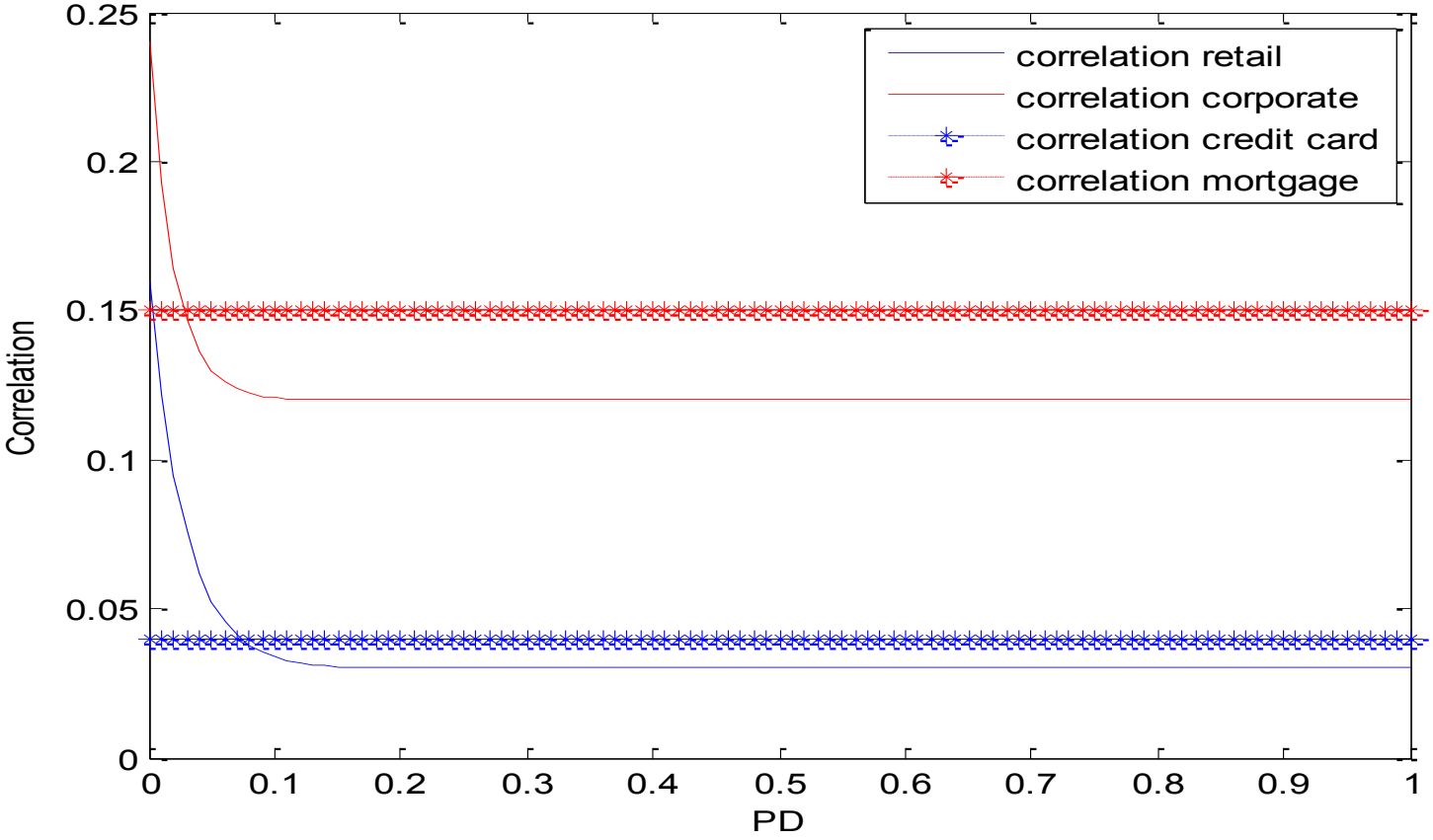
Credit Scoring and Credit Control XIII conference
28th – 30th August 2013

Introduction

- Popular credit risk models use correlation across ***latent variables*** that (supposedly) drive defaults
- Linear correlation (product-moment) is used
- Example: Basel Accord (estimation of capital required to cover unexpected credit losses)
- Different correlations defined for each credit class (e.g. credit card and mortgage)

Motivation (cont'd)

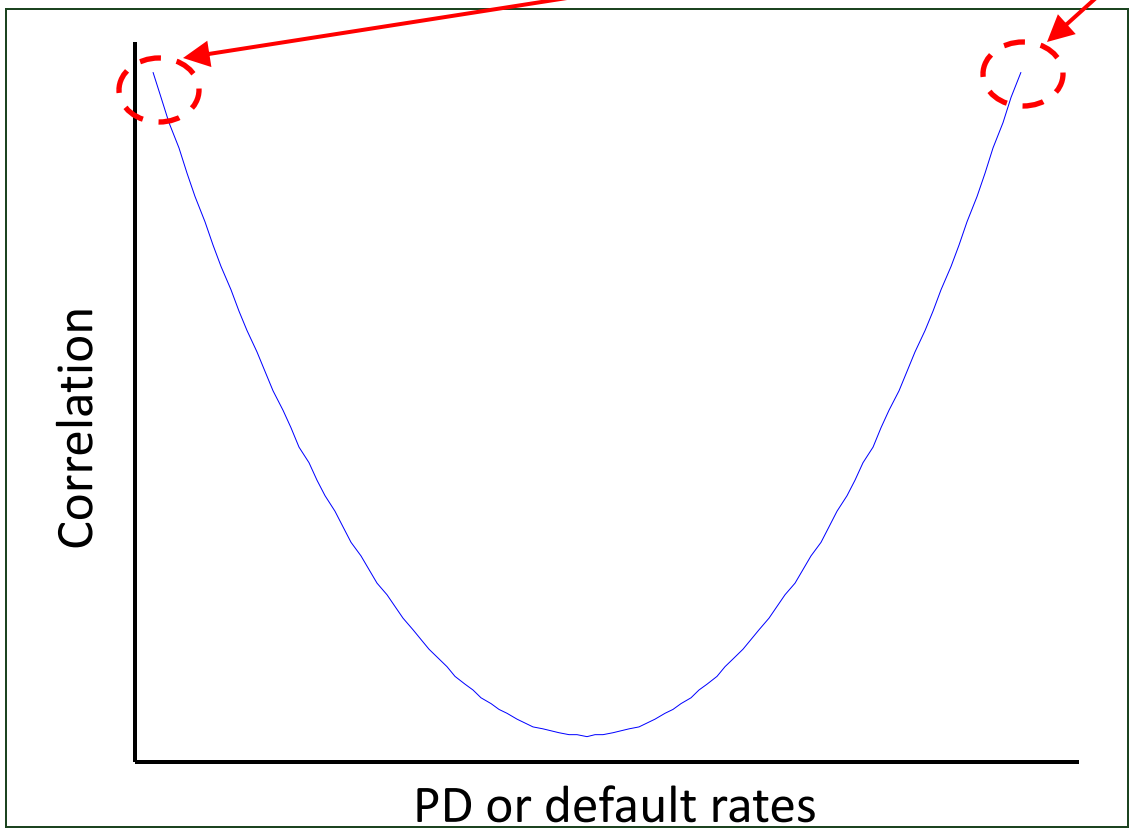
- Limitation: current method (Basel Accords) does ***not*** capture potentially higher correlation *across latent variables* when default rates (or probabilities of default = PD) are either "extremely low" or "extremely high"



Assume that high-risk (high-PD) loans default due to idiosyncratic (specific) factors and therefore present low correlation

Motivation

- We should expect a relationship like the following:



Intuition: highest correlation is related to BOTH "low" and "high" default rates (or PDs) because latent variables (asset returns, e.g.) are more associated **not only when default rates (or PDs) are low but also when they are high (latent variables decrease at the same time regardless of the cause – common or specific).**

Think in terms of portfolios. Interpretation: PD = (expected) default rate in portfolio

Empirical estimates

CRC Credit Research Centre

Basel Accord (II)

Credit Segment	Correlation
Corporate, sovereign, bank exposures, SMEs	between 0.12 and 0.24
High-volatility commercial real estate	between 0.12 and 0.30
Residential mortgages	0.15
Revolving retail credit	0.04
Other retail exposures	between 0.03 and 0.16

The method and the data used were not disclosed

Academic papers

Author(s)	Segment	Technique	Correlation
Rösch (2003 and 2005), Hamerle et al. (2003), Hamerle and Rösch (2006)	Corporate	MLE	0.000462 to 0.02274
Lopez (2004)	Corporate	Min difference models	0.100 to 0.325
Rösch and Scheule (2004)	Retail	Regressions	0.0073 to 0.0102
De Andrade and Thomas (2007)	Retail	Corr across beh scores	0.00047 to 0.00095
Crook and Bellotti (2012)	Credit Cards	MLE	0.00396 to 0.018

An alternative measure: Tetrachoric Correlation

- Measures the correlation between two (artificially) ***dichotomized continuous*** variables that follow a bivariate normal distribution
- It is an estimate of the correlation between the variables if they were not artificially dichotomized

Tetrachoric Correlation: An example

- Relationship between operational condition and hours of operation of machines
- If we observe dichotomized variables: operational condition (as “good” or “bad”) and hours of operation (as “short” or “long”)
- We don’t have data on the actual values, we cannot calculate the linear correlation

Tetrachoric Correlation: An example

- Calculation based on a contingency table:

Hours of operation				
		Short (Y=0)	Long (Y=1)	Total
Operational condition	Good (X=1)	a	b	$a+b$
	Poor (X=0)	c	d	$c+d$
	Total	$a+c$	$b+d$	$a+b+c+d = N$

- Define two variables :

$$\chi_1 = \sqrt{\frac{\pi}{2}} \frac{(a + c) - (b + d)}{N}$$

$$\chi_2 = \sqrt{\frac{\pi}{2}} \frac{(a + b) - (c + d)}{N}$$

Tetrachoric Correlation: An example

- Use χ_1 and χ_2 to find two new variables h and k:

$$h = \chi_1 + \frac{1}{3!}\chi_1^3 + \frac{7}{5!}\chi_1^5 + \frac{127}{7!}\chi_1^7 + \ldots \qquad k = \chi_2 + \frac{1}{3!}\chi_2^3 + \frac{7}{5!}\chi_2^5 + \frac{127}{7!}\chi_2^7 + \ldots$$

- Define $\in$ as:

$$\in = \frac{ad - bc}{N^2\phi(h)\phi(k)}$$

where *a*, *b*, *c*, *d*, and *N* are defined in the contingency table and ϕ is the pdf of the normal distribution

- Find the tetrachoric correlation, r, in:

$$\in = r + \frac{r^2}{2!}hk + \frac{r^3}{3!}(h^2 - 1)(k^2 - 1) + \frac{r^4}{4!}hk(h^2 - 3)(k^2 - 3) + \frac{r^5}{5!}(h^4 - 6h^2 + 3)(k^4 - 6k^2 + 3) + \ldots$$

Since $r \leq 1$ and the denominators $n!$ go to ∞, the terms $r^n/n!$ approach zero.
Therefore, the first few terms are sufficient to determine r.
The tetrachoric coefficient can be estimated in, e.g., SAS® and STATA®.

Tetrachoric correlation: A simplified formula

- As seen before, in:

$$\in = r + \frac{r^2}{2!}hk + \frac{r^3}{3!}(h^2-1)(k^2-1) + \frac{r^4}{4!}hk(h^2-3)(k^2-3) + \frac{r^5}{5!}(h^4-6h^2+3)(k^4-6k^2+3) + \ldots$$

  - Since $r \leq 1$ and the denominators $n!$ go to ∞, the terms $r^n/n!$ approach zero.
  - Therefore, the first few terms are sufficient to determine r.

- From this, we can use the simplified formula restricted to the first term above, r (at the expense of some bias):

$$r = \in = \frac{ad-bc}{N^2\phi(h)\phi(k)}$$

From the previous slide

Tetrachoric correlation in credit risk models

- We observe continuous latent variables (asset returns of obligors) dichotomized into statuses "default" or "non-default" and assumed to be normally distributed
- Thus, *since we do not have data on the actual values of the latent variables, we can use the tetrachoric correlation to estimate their linear correlation*

Tetrachoric correlation in credit risk models

- We need to define the contingency table for this case
- Use the number of *pairs of loans* (joint occurrences) with respect to their statuses (default and non-default)

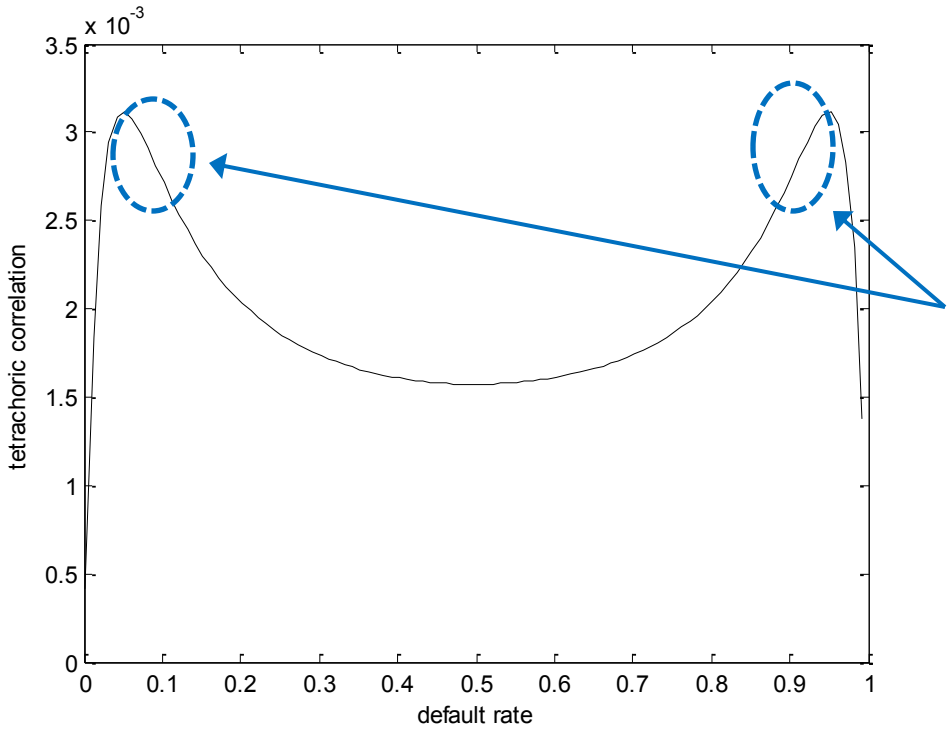
This term, e.g., gives the number of pairs in which both loans defaulted (given the default rate in the period = i.e. number of defaults/total number of loans)

		Default status (loan *i*)	
		default	non-default
Default status (loan *j*)	default	*comb(dr*L,2)*	*[(L – (dr*L)) * (dr*L)]/2*
	non-default	*[(L – (dr*L)) * (dr*L)]/2*	*comb((L – (dr*L)),2)*

where *comb(x,y)* stands for the combination of *x* elements taken *y* a time; *dr* is default rate, and *L* is the number of loans in the portfolio.

Estimates based on the tetrachoric correlation

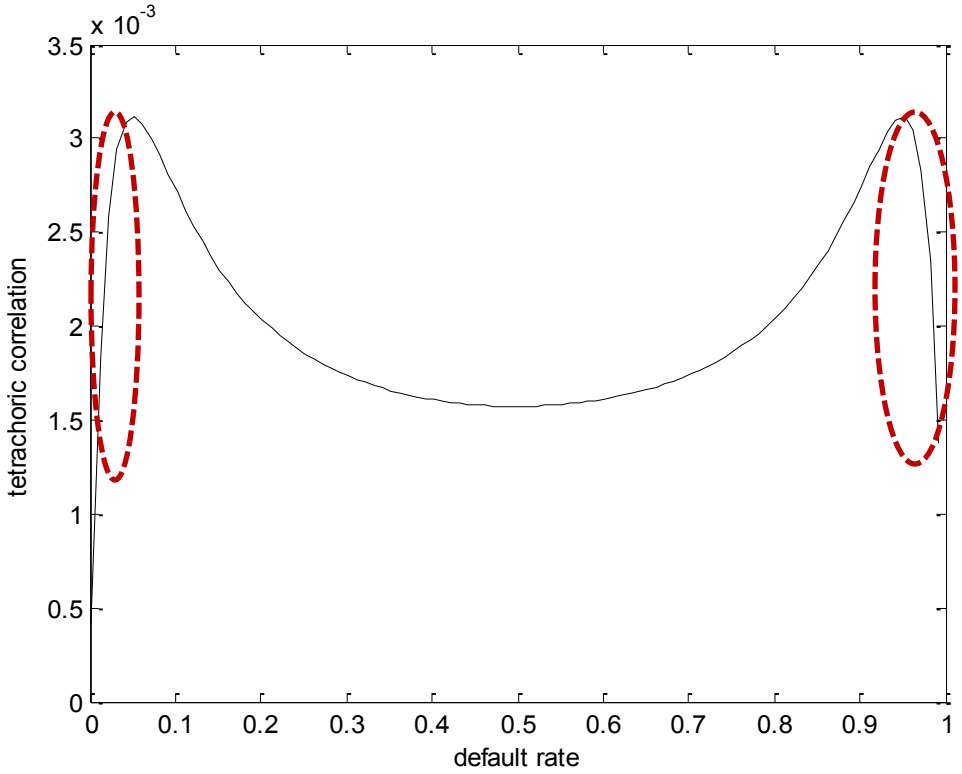
- Estimated asset correlations (i.e. across latent variables) for portfolios composed of 100, 1000, and 10000 loans (conditional on several default rates). Example: 1000 loans



- Lowest correlation = 0.000412 (for dr = 0.02)
- Highest correlation = 0.003113 (for dr = 0.05 and 0.95)
- Intuition: for either very low or very high default rates, asset returns of obligors tend to move together (debtors are in similar situation), so correlation is higher
- For intermediate values of default rates, correlations are lower
- However …

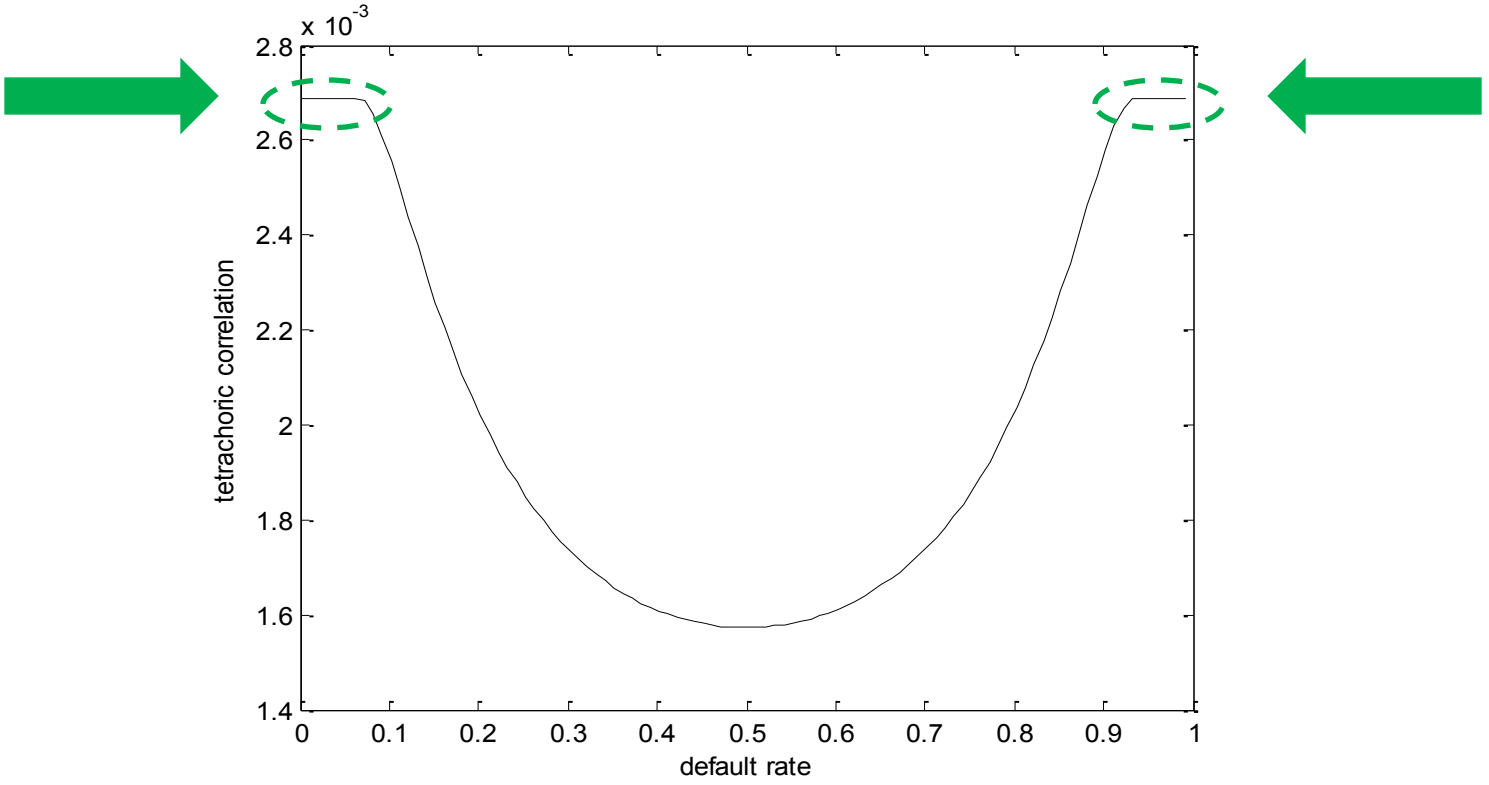
Estimates based on the tetrachoric correlation

- However, cannot explain the correlations found for the extreme values of default rates, i.e., why do they decrease?



Estimates based on the tetrachoric correlation

- Therefore, I propose an artificial adjustment (until a solution is found)



Limitations of this paper

- What I am *not* doing in this paper:
  - I am not relaxing the assumption of normality (regarding the latent variables in these models)
  - I am not relaxing the assumption of homogenous correlation (equal value for all pairs) in a particular portfolio

Conclusions

- Theoretically, the tetrachoric correlation is more compatible (than the linear correlation) with traditional credit risk models (based on latent variables)
- More realistic since it reflects the current situation of loan portfolios such that “low” or “high” default rates indicate “higher” correlation
  - in contrast to Basel which specifies a monotonic decreasing relationship between correlation and probability of default (i.e. default rates)

Conclusions

- Instead of focusing on credit segments (as Basel does), I focus on default rates
- Can be used as time-varying model (updated according to the most recent default rates)
- Some work is still needed to estimate the correlation regarding extreme values of default rates