Estimating Long-Run PD, Asset Correlation, and Portfolio Level PD by Vasicek Models

Yang, Bill Huajian

10 July 2013

ESTIMATING LONG-RUN PD, ASSET CORRELATION, AND PORTFOLIO LEVEL PD BY VASICEK MODELS*

(Pre-typeset version)
(Final version is published in "Journal of Risk Model Validation", Vol.7/No.4, 2013)

BILL HUAJIAN YANG

Abstract

In this paper, we propose a Vasicek-type of models for estimating portfolio level probability of default (PD). With these Vasicek models, asset correlation and long-run PD for a risk homogenous portfolio both have analytical solutions, longer external time series for market and macroeconomic variables can be included, and the traditional asymptotic maximum likelihood approach can be shown to be equivalent to least square regression, which greatly simplifies parameter estimation. The analytical formula for long-run PD, for example, explicitly quantifies the contribution of uncertainty to an increase of long-run PD. We recommend the bootstrap approach to addressing the serial correlation issue for a time series sample. To validate the proposed models, we estimate the asset correlations for 13 industry sectors using corporate annual default rates from S&P for years 1981-2011, and long-run PD and asset correlation for a US commercial portfolio, using US delinquent rate for commercial and industry loans from US Federal Reserve.

Keywords: Portfolio level PD, long-run PD, asset correlation, time series, serial correlation, bootstrapping, binomial distribution, maximum likelihood, least square regression, Vasicek model

1. Introduction

For a risk homogeneous portfolio, the long-run PD (LRPD) refers to the expected value of portfolio default rate in one-year horizon. It reflects the bank's long-term view of portfolio default risk, and is usually the target PD a rating model calibrates to. Long-run PD and asset correlation both are key components for assessments of capital requirements ([4], [6], [7], [9], [10], [12], [13], [19]).

Due to the insufficient internal portfolio default data, estimation of long-run PD and asset correlation proves to be difficult ([4], [7], [10], [12], [13], [19]). Traditional asset correlation estimation methodologies include the binomial maximum likelihood approach for observed default counts and the asymptotic maximum likelihood approach where observed default rates are equated to the portfolio level PD.

Let n denote the size of the portfolio, and k the number of defaults in one-year horizon. Portfolio default rate at the horizon is given by $r = k/n$. We assume that the default count k follows a binomial distribution, given the event probability $p(s)$ dictated by a latent effect s. We call this $p(s)$ the portfolio level PD given s, and write p for $p(s)$ when it causes no ambiguity. We can think of $p(s)$ as the asymptotic portfolio default rate, when portfolio size is sufficiently large ([7]).

Let Φ denote the cumulative distribution for a standard normal variable. Recall that a random variable y follows a Vasicek distribution if $\Phi^{-1}(y)$ is normal ([12, p52]). Thus a Vasicek distribution is determined by the mean and standard deviation of $\Phi^{-1}(y)$.

* Bill Huajian Yang, Ph. D in mathematics, Wholesale Credit Methodologies, Bank of Montreal, Canada. Mail address: 51st floor, First Canadian Place, 100 King Street West, Toronto, Canada, M5X 2A1

The views expressed in this article are not necessarily those of Bank of Montreal or any of its affiliates.

Please direct any comments to bill.yang@bmo.com, phone 416-643-1922

We propose two types of Vasicek models for estimating portfolio level PD:

(I) $p(s) = \Phi(a + bs),\ s \sim N(0,1)$

(II) $p(s) = \Phi(a + \sum_{i=1}^{m} b_i s_i + cs),\ s \sim N(0, 1)$

Here $s_1, s_2, \ldots, s_m$ are market factors or macroeconomic variables, subjected to an appropriate transformation by Φ^{-1} when necessary. The latent random effect s is independent of $s_1, s_2, \ldots, s_m$. It represents the model residual in presence of market or macroeconomic factors $s_1, s_2, \ldots, s_m$, including for example, the effects and dynamics account for default contagion ([5]). Modeling portfolio level PD rather than default rate eliminates the effects of portfolio size.

Clearly, type I models are the simplest form of type II models. Theoretical results shown under the type I model framework can be applied to type II models under the assumption that $s_1, s_2, \ldots, s_m$ are normal. Type II models are particularly useful when internal portfolio historical default data is limited, but longer external market and macroeconomic variables are available. The use of longer market time series improves the quality of parameter estimation.

The proposed models are a type of generalized linear models, targeting portfolio level PD. As shown later in section 2, the models can be reformulated as a type of probit models. Thus portfolio level PD can be interpreted as the probability of a binary credit event conditional on market or macroeconomic factors and a latent random effect. The models are slightly different from Merton models ([11], [17]) of the form $\Phi(z)$, where z is normalized and interpreted as the normalized asset value for a borrower (see section 3).

As shown in later sections, the advantages of the proposed Vasicek models for portfolio level PD include:

(a) Asset correlation and long-run PD for a risk homogenous portfolio both have analytical solutions under type I model framework (Proposition 3.1)
(b) Asymptotic maximum likelihood approach is equivalent to least square regression (Theorem 4.2)
(c) Longer external time series of market and macroeconomic variables can be included by a type II model (see section 6 for examples)

This paper is organized as follows. Some key theoretical results for the proposed models are shown in section 2. In section 3, we derive the analytical formulas for asset correlation and long-run PD for a risk homogenous portfolio. In section 4, we review the traditional parameter estimation methodologies, show the equivalence between the asymptotic maximum likelihood and least square regression, and propose an asymptotic least square approach with a variance correction. In section 5, we discuss the serial correlation issues for a time series sample, and propose a bootstrap approach as a fix. We validate the proposed models and approaches in section 6 by estimating: (a) the asset correlations for 13 industry sectors using corporate annual default rates posted by S&P for years 1981-2011; (b) the long-run PD and the asset correlations for a US commercial portfolio, using external US Federal Reserve delinquent rates for commercial and industry loans as a market factor.

The author thanks the referees for valuable comments, particularly the comments for the example in section 6.1; and Clovis Sukam for many pleasant and valuable conversations.

2. Some Basic Results for the Proposed Vasicek Models

Recall that the default count *k* is assumed to follow a binomial distribution, given the event probability $p(s)$, dictated by a latent effect *s*. The portfolio level PD (i.e., $p(s)$) differs essentially from the default rate *r*. We can think of a realization of portfolio default rate *r* as consisting of two random processes: The first or inner process generates $p(s)$, which is governed by a latent effect *s*, the second or outer process generates *k* (thus *r*) following the binomial distribution with event probability $p(s)$. Therefore, by law of total variance, the variance for *r* is always larger than that of $p(s)$.

The assumption of binomial distribution for default count *k* conditional on $p(s)$ implies that obligors in the portfolio default independently conditional on $p(s)$ (which does not mean unconditional independence).

The following lemma on expected value of $\Phi(a+bs)$ is useful for later discussions (see Appendix for a proof).

Lemma 2.1. ([16, p47]) Let $s \sim N(0,1)$ be standard normal. Then

$$E(\Phi(a+bs)) = \Phi(a/\sqrt{1+b^2})$$

Given a value of the sum $(a+\sum_{i=1}^{m} a_b s_i)$, we can drop off the random effect *s* for a type II model using Lemma 2.1, and calculate the model predicted portfolio level PD as in the corollary below.

Corollary 2.2. $E[\Phi(a+\sum_{i=1}^{m} b_i s_i + cs) \mid s_1, s_2, ..., s_m] = \Phi[(a+\sum_{i=1}^{m} a_b s_i)/\sqrt{1+c^2}]$

Let $E(r), v(r)$ denote the expected value and variance of portfolio default rate *r,* and p_0, v_0 the expected value and variance of $p(s)$, the portfolio level PD. Let *D* be a Bernoulli trial with event probability p_0. Then $E(D) = p_0$ and $v(D) = p_0(1-p_0)$. Proposition 2.3 (c) below is to be used in section 4.3 for a variance correction to the asymptotic least square approach. The quantity v_0 will be further discussed by the joint default probability in Proposition 3.3.

Proposition 2.3. (a) $E(r) = p_0$

(b) $v(D) = p_0(1-p_0) \geq v(r) \geq v_0$

(c) $v_0 = v(r) - [p_0(1-p_0) - v(r)]/(n-1)$

Proof. The expected value of a binomial distribution, given the event probability *p* (i.e., $p(s)$), equals to np. Thus the expected value of *r*, conditional on *p*, is *p*. Therefore

$$E(r) = E[E(r \mid p)] = E(p) = p_0$$

and we have (a). Next, conditional on event probability *p*, the variance of a Bernoulli trial *D* is $p(1-p)$. By law of total variance, we have:

$$v(D) = p_0(1-p_0) = E[v(D \mid p)] + E(p-p_0)^2 = E[p(1-p)] + v_0 \quad (1)$$

Conditional on *p,* the variance of *r* is $p(1-p)/n$, we thus have:

$$v(r) = E(r - p_0)^2 = E[E(r-p)^2 \mid p] + E(p - p_0)^2$$
$$= E[p(1-p)/n] + v_0 \qquad (2)$$
$$= [p_0(1-p_0) - v_0]/n + v_0 \qquad (3)$$
$$= p_0(1-p_0)/n + (n-1)v_0/n$$

where (3) follows from (1). Because $p(1-p) \geq p(1-p)/n$, we have (b) by (1) and (2). Statement (c) follows from (3). □

We end this section by mentioning that both types I and II models are a type of probit models, which means $p(s)$ is the probability of a binary credit event, given s and $s_1, s_2, \ldots, s_m$. For example, for a type II model, we have:

$$p(s) = \Phi(a + \sum_{i=1}^{m} b_i s_i + cs)$$
$$= P(\varepsilon \leq a + \sum_{i=1}^{m} b_i s_i + cs), \quad \varepsilon \sim N(0,1)$$
$$= P(D = 1 \mid s, s_1, s_2, \ldots, s_m)$$

where ε can be interpreted as a variable measuring a type of credit risk for the portfolio, and the event variable D is defined to have value 1 if $\varepsilon \leq a + \sum_{i=1}^{m} b_i s_i + cs$, and is 0 otherwise.

3. Analytical Formulas for LRPD and Asset Correlation for a Homogeneous Portfolio

Under the Merton model framework ([11], [17]), obligor's default risk is driven by a normalized latent variable z. A default event for an obligor occurs as soon as z falls below a threshold value called default point. Different obligors may have different default points. In the case when portfolio default risk is homogenous between obligors, we can assume that obligors have the same default point d_0. One factor Merton for a risk homogenous portfolio assumes that z splits into two parts:

$$z = \sqrt{\rho}\, s + \sqrt{1-\rho}\,\varepsilon, \; 0 \leq \rho \leq 1 \qquad (4)$$

where s and ε are independent random variables, both standard normal, with s the systemic risk, common to all obligors in the portfolio, and ε the idiosyncratic risk. The quantity ρ is called the asset correlation between obligors. Portfolio level PD given s is given by

$$p(s) = P(\sqrt{\rho}\, s + \sqrt{1-\rho}\,\varepsilon \leq d_0)$$
$$= P(\varepsilon \leq (d_0 - \sqrt{\rho}\, s)/\sqrt{1-\rho})$$
$$= \Phi((d_0 - \sqrt{\rho}\, s)/\sqrt{1-\rho})$$

Therefore Merton models for a risk homogenous portfolio are one-parameter (i.e. ρ) type I models with default point given and specified. By Lemma 2.1, we have

$$E(p(s)) = E[\Phi((d_0 - \sqrt{\rho}\, s)/\sqrt{1-\rho})\;] = \Phi(d_0) \qquad (5)$$

For a risk homogenous portfolio, both long-run PD and asset correlation have analytical solutions under the type I model framework:

Proposition 3.1. Given a type I model $\Phi(a+bs)$ for the portfolio level PD, one has:

(a) Expected portfolio level PD $=\Phi(a/\sqrt{1+b^2})$, mode and median PD $=\Phi(a)$
(b) Default point $=a/\sqrt{1+b^2}$
(c) Default implied asset correlation $\rho=b^2/(1+b^2)$.

Proof. Obviously, the mode and median PD is given by $\Phi(a)$. The formula for expected PD follows from Lemma 2.1. By (5) and (a), we have (b). Statement (c) follows from the fact that the normalization $(\varepsilon+bs)/\sqrt{1+b^2}$ splits into two parts as in (4), with the coefficient of s given by $b/\sqrt{1+b^2}$.□

By Proposition 3.1 (a), when $a<0$ (i.e., when expected portfolio level PD is less than 50%), the expected portfolio level PD increases as uncertainty (variance b^2) increases, and is in general larger than the mode/median PD for a type I model. The mode PD is considered by Canadian financial institution regulator, the Office of the Superintendent of Financial Institution (OSFI), to serve as the long-run PD ([13], [14]). However, under the type I model, this mode/median PD is lower than the expected portfolio level PD. Therefore, it is a biased estimator for the long-run PD. For a risk homogenous portfolio, let $LRPD_{\exp}$ and $LRPD_{\mathrm{mod}}$ denote the expected portfolio level PD and the mode/median PD respectively. We thus have:

Proposition 3.2. Under the type I model framework, one has $LRPD_{\mathrm{mod}} < LRPD_{\exp}$ if the expected portfolio level PD is less than 50%.

To end this section, we summarize in the next proposition the relationships between the asset correlation ρ, the pair-wise default correlation ρ^D, and the mean and variance of portfolio level PD (i.e., p_0 and v_0). Let *JDP* denote the joint default probability between two borrowers, which is given by:

$$JDP(p_0,\rho)=N_2(\Phi^{-1}(p_0),\Phi^{-1}(p_0),\rho)$$

where $N_2(x_1,x_2,\rho)$ denotes the cumulative distribution for bivariate standard normal variables (x_1,x_2) with asset correlation ρ as the correlation between variables x_1 and x_2.

Proposition 3.3. For a risk homogenous portfolio, one has:
(a) ([16, p48]) $v_0=JDP(p_0,\rho)-(p_0)^2$
(b) ([19, pp7-10]) $\rho^D=(JDP(p_0,\rho)-(p_0)^2)/[p_0(1-p_0]=v_0/[p_0(1-p_0]$

4. Model Parameter Estimation Methodologies

Traditional asset correlation estimation methodologies include:

(a) Binomial maximum likelihood
(b) Asymptotic maximum likelihood

Binomial maximum likelihood maximizes the likelihood of observing portfolio default counts, assuming default count follows a binomial distribution given the event probability. Asymptotic maximum likelihood approaches equate the observed portfolio default rate to the portfolio level PD ($p(s)$), and maximizes the likelihood of observing portfolio default rates. We show in this section the equivalence between the asymptotic maximum likelihood approach and the least square regression for a type II model.

4.1. The Likelihood of Portfolio Level PD and Least Square Regression

The following proposition on likelihood of portfolio level PD is important, with which we can show the equivalence between asymptotic maximum likelihood approach and least square regression. Proposition 4.1 (a) holds in general for any cumulative distribution $\Phi(x)$ and density $\phi(x)$, not just for normal distributions (see appendix for a proof).

Proposition 4.1. Let $z = \Phi^{-1}(p)$ where $p = \Phi(a + bs)$ and $s \sim N(0, 1)$.

(a) The likelihood of p is given by:

$$d(p) = \phi((z - a)/b)/(b\phi(z))$$

(b) The negative log likelihood of p is given by

$$[(\Phi^{-1}(p) - a)^2/(2b^2)] + \ln(b) - (\Phi^{-1}(p))^2/2$$

Suppose $S = \{(s_{1k}, s_{2k}, ..., s_{mk}, p_k\}$ is a given sample with size N, and p_k as the portfolio level PD at time k, where $s_1, s_2, ..., s_m$ are market or macroeconomic variables, subjected to an appropriate transformation by Φ^{-1} when necessary. Set $z_k = \Phi^{-1}(p_k)$. Assume a type II model for the portfolio level PD:

$$p = \Phi(a + \sum_{i=1}^{m} b_i s_i + cs), \quad s \sim N(0, 1) \qquad (6)$$

The following theorem shows the equivalence between the asymptotic maximum likelihood approach and least square regression (see appendix for a proof).

Theorem 4.2. For a type II model, the maximum likelihood approach is equivalent to least square regression minimizing the sum-square of errors:

$$\sum_{k=1}^{N} [z_k - (a + \sum_{i=1}^{m} b_i s_{ik})]^2 \qquad (7)$$

and c is estimated as the standard deviation of the errors: $z_k - (a + \sum_{i=1}^{m} b_i s_{ik})$.

For example, for a type I model, a special case of the type II models, parameters a and b can be estimated by Theorem 4.2 as:

$$a = (z_1 + z_2 + ... + z_N)/N, \quad b^2 = [(z_1 - a)^2 + (z_2 - a)^2 + ... + (z_N - a)^2]/N$$

Thus we have an estimator for long-run PD of the form $\Phi(a/\sqrt{1+b^2})$ by Proposition 3.1, which is similar to the estimator given by Miu and Ozdemir ([13, (13b)]), using the relations $1 + b^2 = 1/(1-\rho)$ and $b = \sqrt{\rho/(1-\rho)}$. We also have an estimator for asset correlation of the form $\rho = b^2/(1+b^2)$, similar to the estimator given by Meyer ([12, 4.3.2]).

Here we assume that p_k is the portfolio level PD, not the portfolio default rate. In the latter case, the above estimators for long-run PD and asset correlation could be biased when portfolio size is not sufficiently large. We will deal with this issue and propose in section 4.3 an asymptotic methodology with a variance correction.

4.2. Binomial Likelihood Approaches

Let $S = \{(s_{1i}, s_{2i}, ..., s_{mi}, k_i, n_i\}$ be a sample with size *N*, where $s_1, s_2, ..., s_m$ are market or macroeconomic variables, and n_i, k_i are the numbers of obligors and one-year horizon defaults respectively at time index *i*. Given the event probability $p = p(s)$, the likelihood of observing *k* defaults for a portfolio with *n* obligors is:

$$\binom{n}{k} p(s)^k (1 - p(s))^{n-k}$$

Its expected value with respect to *s* gives the unconditional likelihood:

$$bin(k,n) = \binom{n}{k} \int_{-\infty}^{+\infty} p(s)^k (1 - p(s))^{n-k} \phi(s) ds$$

The negative natural logarithmic likelihood is the sum ([7], [10]):

$$-\log L = -\sum_{i=1}^{N} \ln(bin(k_i, n_i))$$

With the maximum likelihood approach for type II models, we are required to find the model parameters by minimizing -log L. SAS non-linear mixed procedure (NLMIXED, [18]) provides a tool for fitting this type of models maximizing the binomial likelihood.

4.3. Asymptotic Least Square Approaches

Let $S = \{(s_{1k}, s_{2k}, ..., s_{mk}, r_k, n_k\}$ be a sample with size *N*, where r_k denotes the default rate at time *k*. We consider two asymptotic approaches.

Case I. Asymptotic (no variance correction)

In this case, we equate the observed default rate r_k to portfolio level PD, and transform r_k to $z_k = \Phi^{-1}(r_k)$. We then estimate the parameters for model (6) through least square regression by Theorem 4.2.

Note that the transformation $\Phi^{-1}(r_k)$ requires $0 < r_k < 1$. Appropriate flooring or capping for the default rate is required when 0 or 1 appears in the time series, otherwise, extreme values of $\Phi^{-1}(r_k)$ will inflate the estimation. Though the default rate can be zero for a small sample, we don't expect zero PD. Besides, if we didn't observe any default when size is small, it does not mean we would not find one when size is increased.

Case II. Asymptotic with a variance correction

Because equating default rate to portfolio level PD exaggerates the variance of portfolio level PD by Proposition 2.3 (b), leading to an overestimate of asset correlation, we propose a variance correction as follows:

Correction for the variance of portfolio level PD:

(a) Assume a constant size *n* for the portfolio over time. First, estimate p_0 as the simple average of sample default rates, and $v(r)$ as the sample variance, then compute by Proposition 2.3 (c) the variance of portfolio level PD as:

$$v_0 = v(r) - [p_0(1 - p_0) - v(r)]/(n - 1)$$

(b) Let $\bar{r}$ denote the sample average of all r_k, and $w = \sqrt{v_0 / v(r)}$. Replace r_k by rr_k:

$$rr_k = \bar{r} + (r_k - \bar{r})w$$

The rest is the same

Note that $p_0(1 - p_0) - v(r) = (r_1 + r_2 + ... + r_m)/m - (r_1^{\ 2} + r_2^{\ 2} + ... + r_m^{\ 2})/m > 0$ unless $r_k = 0$ or 1 for all k. We thus have $w = \sqrt{v_0 / v(r)} < 1$ and $rr_k > 0$ in general. This correction has the advantage of transforming extreme values of 0 and 1 to other regular values between 0 and 1, which would have been an issue for the traditional asymptotic approach with no variance correction. More importantly, the sample variance of rr_k is now adjusted to the sample variance v_0 of portfolio level PD (i.e. $p(s)$).

5. The Serial Correlation and Bootstrap Methodologies

Traditionally, historical default rate time series sample is used for model fitting. However, serial correlation for a times series sample is in general significant, and is an issue for parameter estimation ([13], [15, pp.159-175]).

For a default rate time series, the serial correlation is in general positive. This positive serial correlation causes the sample variance of errors $\sum_{k=1}^{N} (z_k - a)^2 / N$ to overestimate the parameter b^2 under the type I model framework. By Proposition 3.1, this results in an overestimate for the asset correlation, and an underestimate for the long-run PD (when $a < 0$, i.e., when the expected portfolio level PD is less than 50%).

A model describes the joint distribution between the target and explanatory variables. Given a modeling sample, independence between data points is generally expected and required.

Instead of fitting the model directly on the time series sample, we propose a bootstrap approach: Generate, say 1000, bootstrap samples each is of the same size as the original time series sample, and is sampled randomly from the original time series sample with replacement ([8, section 8.2]). Fit a model over each bootstrap sample and estimate the long-run PD and asset correlation. The final long-run PD and asset correlation are estimated by averaging all the estimates from the bootstrap samples. This is analogous to the bagging technique ([3], [8]). Confidence bounds can be calculated when the number of bootstrap samples is sufficiently large.

We thus propose the following bootstrap steps for estimating long-run PD and asset correlation:

(a) Generate B (sufficiently large, say 1000) bootstrap samples using the time series sample.

(b) For each bootstrap sample, fit a type II model:

$$p(s) = \Phi(a + \sum_{i=1}^{m} b_i s_i + cs), \ s \sim N(0, 1)$$

(c) Let $u = \sum_{i=1}^{m} b_i s_i$ and $v = \sum_{i=1}^{m} b_i s_i + cs$. Assume that $s_1, s_2, ..., s_m$ are normal. Estimate the mean m_1 and standard deviation σ_1 for u over the bootstrap sample, and calculate the mean and variance of v by m_1 and $\sigma_1^{\ 2} + c^2$

(d) Estimate by Proposition 3.1 the asset correlation, long-run PD, mode, and median PD as:

$$\rho = (\sigma_1^{\ 2} + c^2)/(1 + \sigma_1^{\ 2} + c^2), \quad p_0 = \Phi[(a + m_1)/\sqrt{1 + \sigma_1^{\ 2} + c^2}\,]$$

Mode and median $PD = \Phi(a + m_1)$

6. Empirical Examples

6.1. Asset Correlations for Industry Sectors by S&P Corporate Annual Default Rates

In this section, we estimate asset correlations for 13 industry sectors based on corporate annual default rate time series posted by S&P for years 1981-2011. We follow the steps proposed in section 5 and bootstrap 1000 times. Each time we fit a type II model for each sector using the yearly default rate over all sectors as a common market factor:

$$p(s) = \Phi(a + bs_1 + cs), \quad s \sim N(0, 1)$$

where $s_1 = \Phi^{-1}(\text{yearly default rate over all sectors})$

Given the default rate time series for a sector, we assume that the yearly number of firms rated for the sector is constant across years, and is equal to the average of numbers of firms rated for the sector across years in the sample. Table 1 shows the results by three approaches: binomial maximum likelihood, asymptotic least square with ("Asymptotic Adj") and without ("Asymptotic No Adj") variance correction. The last column shows the 95% percentile upper bound for the bootstrap estimate under the asymptotic approach with a variance correction. These results are comparable to the results by Demey ([7]), which were based on S&P annual corporate default rates for years 1981-2002. Results by bootstrap method are compared to the traditional method, where the original time series sample is used directly.

Table 1. Asset correlation for industry sectors

		Traditional method			Bootstrap method			
			Asymptotic			Asymptotic		
Industry Sector	Avg DR	Binomial	Adj	No Adj	Binomial	Adj	No Adj	P95 (Adj)
Aerospace/automobile	2.3%	9.9%	10.1%	13.1%	8.9%	9.0%	12.0%	13.3%
Consumer/service	2.4%	7.0%	7.1%	10.4%	5.8%	6.2%	9.2%	8.8%
Energy/natural resources	2.1%	10.2%	12.4%	17.5%	9.3%	11.2%	16.7%	15.9%
Financial institutions	1.2%	14.1%	12.7%	14.1%	12.1%	10.8%	13.4%	17.7%
Forest/building products	2.6%	15.1%	16.1%	19.3%	13.6%	14.7%	18.3%	20.5%
Health	3.2%	23.7%	25.5%	26.5%	19.7%	20.5%	24.6%	37.9%
High tech	2.3%	13.1%	18.7%	21.8%	11.8%	15.8%	21.1%	26.6%
Insurance	0.9%	3.0%	5.7%	9.3%	3.0%	5.0%	9.2%	9.5%
Leisure time/media	3.4%	15.1%	14.7%	18.3%	13.5%	13.3%	17.1%	19.0%
Real estate	1.7%	13.8%	19.0%	22.4%	12.6%	17.8%	21.6%	23.4%
Telecoms	2.5%	15.2%	20.7%	24.9%	13.6%	18.7%	23.6%	26.9%
Transportation	2.1%	3.2%	5.9%	16.5%	3.4%	5.7%	15.9%	8.9%
Utilities	0.7%	2.2%	4.6%	6.9%	2.4%	4.0%	6.7%	8.9%

The results show: (a) Overall, results by asymptotic approach with a variance correction are fairly consistent with the results by binomial approach. (b) Estimates by bootstrap method are generally lower than the traditional method as expected (see section 5), and are significant for some cases (e.g., Health). This means the serial correlation would have inflated the estimates should one use the traditional method. (c) Asset correlations are mostly below or around 15%, except for Health, High Tech, Real Estate, and Telecoms, where high variances are found from the observed default rate time series (default correlation and asset correlation, are positively correlated with the variance of portfolio level PD, by Proposition 3.3 (b)). (d) The results of asymptotic approach with a variance correction are always lower than results with no correction. (e) The 95% percentile upper bound for the estimate under the asymptotic approach with a variance correction is in general not too significantly higher than the estimate itself, except for Health, High Tech, and Telecoms, which are again due to the high variances of the observed default rates.

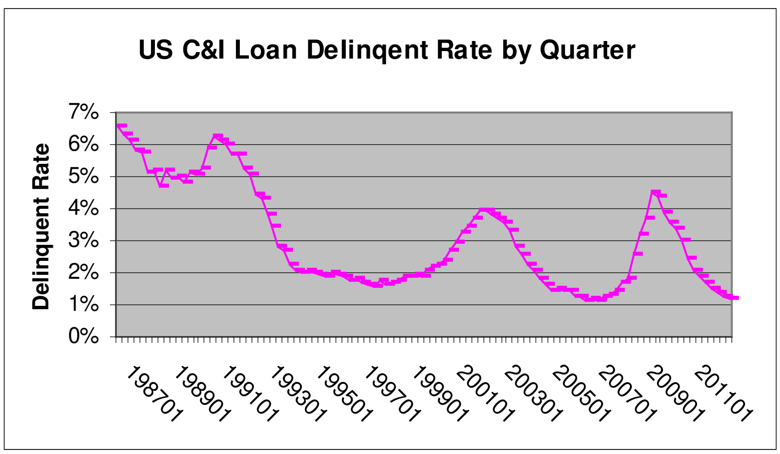
We should be cautious when interpreting the results by asymptotic approach with no adjustment. With this approach, default rate has to be floored appropriately, as pointed out in section 4.3. We floor the default rate at 0.2% when default rate is 0. This could have inflated the estimates for some sectors, such as Energy, High Tech, Telecoms, and Transportation. For this reason, estimate by asymptotic approach with a variance correction or by binomial approach is preferred.

With this S&P data, the default rate time series is split by year. Splitting by 6 months or by quarter will give more data points, which will definitely be helpful for parameter estimation. This is what we do for the example in next section.

6.2. Long-Run PD and Asset Correlation for a Commercial Portfolio

In this section, we estimate the asset correlation, long-run PD for a US commercial portfolio, where historical one-year default rates for the portfolio are available for each quarter between 2006 and 2012.

The market variable we use is the delinquent rate for commercial and industry loans (no seasonal adjustment), posted by US Federal Reserve, available since 1987. The chart below depicts the US delinquency since 1987. There have been two cycles since 1995:



Based on historical default rate data, portfolio default rate responds to US delinquent rate by a lag of two quarters. We thus shift up the US delinquent rate by two quarters to line up with the internal portfolio default rate for modeling purpose.

Again, we follow the steps proposed in section 5 and bootstrap for 1000 times. Each time we fit a type II Vasicek model of the form

$$p(s) = \Phi(a + bs_1 + cs), \quad s \sim N(0, 1)$$

where $s_1 = \Phi^{-1}(US\ delinquent\ rate)$

Table 2 below shows the asset correlation, median/mode PD, and long-run PD by three approaches. The results show, bootstrap method estimates slightly lower asset correlations, but slightly higher long-run PDs. This is expected as explained in section 5. We calculate the bootstrap 95% percentile value of the long-run PD as 3.47%, under the asymptotic approach with a variance correction. This can be used as an upper confidence bound.

Table 2. Long-run PD and Asset Correlation on a commercial portfolio using two cycles of US delinquent rates since 199501

		Traditional method			Bootstrap method		
Estimating Methodology	Avg DR	Median Value PD	Expected Value PD	Asset Correlation	Median Value PD	Expected Value PD	Asset Correlation
Binomial	3.01%	2.73%	3.07%	5.16%	2.79%	3.11%	4.7%
Asym Adj	3.01%	2.76%	3.13%	5.34%	2.82%	3.16%	4.9%
Asym	3.01%	2.78%	3.14%	5.81%	2.85%	3.16%	5.3%

Conclusion. With the proposed Vasicek models for portfolio level PD, asset correlation and long-run PD for a risk homogenous portfolio both have analytical solutions, parameters can be estimated through least square regression, which is simple and easy to implement. These Vasicek models have the advantages of incorporating longer external market and macroeconomic variables, improving the quality of parameter estimation in contrast to using only a shorter period of internal default data. The proposed bootstrap approach corrects the bias of parameter estimates by traditional method due to serial correlation. We believe that the proposed models, the bootstrap technique, and the asymptotic least square approach with a variance correction are potentially a good tool for assessments of long-run PD, asset correlation, and portfolio level PD.

REFERENCES

[1] Basel Committee on Banking Supervision (2005). An Explanatory Note on the Basel II IRB Risk Weight Functions, July 2005. http://www.bis.org/bcbs/irbriskweight.pdf

[2] Blaschke, W., Jones, M., Majnoni, G., and Peria, S. (2001). Stress Testing of Financial Systems: An Overview of Issues, Methodologies, and FSAP Experiences, IMF working paper, June 2001

[3] Breiman, L. (1996). Bagging Predictors. Machine Learning 24: 123-140 http://www.cs.utsa.edu/~bylander/cs6243/breiman96bagging.pdf

[4] Chernih, A., Vanduffel, S., and Henald, L. (2006). Asset correlations: a literature review and analysis of the impact of dependent loss given defaults http://www.econ.kuleuven.be/insurance/pdfs/CVH-AssetCorrelations_v12.pdf

[5] Das, S., Duffie, D.,Kapadia, N., Saita, L. (2007) Common Failings: How Corporate Defaults are Correlated. Journal of Finance 62 (1), 93-117 http://www.q-group.org/archives_folder/pdf/Paper-Das.pdf

[6] De Servigny, A., Renault, O. (2003). Default Correlation Evidence. Risk, July 2003, 90-94.

[7] Demey, P., Jouanin, J., Roget, C, and Roncalli, T. (2004). Maximum likelihood estimate of default correlations, Risk, November 2004 http://thierry-roncalli.com/download/risk-mledc.pdf

[8] Friedman, J., Hastie, T., and Tibshirani, R. (2008). The Elements of Statistical Learning, 2nd edition, Springer

[9] Frye, J. (2008). Correlation and asset correlation in the structural portfolio model. The Journal of Credit Risk, Volume 4 (2), Summer 2008 http://www.risk.net/digital_assets/4538/v4n2a4.pdf

[10] Gordy, M., Heitfield, E. (2002). Estimating default correlations from short panels of credit rating performance data. Federal Reserve Board Working paper, January 2002

[11] Merton, R. (1974). On the pricing of corporate debt: the risk structure of interest rates. Journal of Finance, Volume 29 (2), 449-470 http://www.jstor.org/discover/10.2307 /2978814?uid=2129&uid=2&uid=70&uid=4&sid=21102236625247

[12] Meyer, C. (2009). Estimation of intra-sector asset correlations. The Journal of Risk Model

validation, Volume 3 (3), Fall 2009
http://www.risk.net/digital_assets/5021/jrm_v3n3a3.pdf
[13] Miu, P., Ozdemir, B. (2008). Estimating and validating long-run probability of default with respect to Basel II requirements. The Journal of Risk Model validation, Volume 2/Number 2, 3-41. http://www.risk.net/digital_assets/5008/jrm_v2n2a1.pdf
[14] OSFI (2004). Risk quantification of IRB systems at IRB banks: Appendix – A conservative Estimate of a long-term average PD by a hypothetical bank, December 2004
[15] Pindyck, R. S., Rubinfeld, D. L. (1998). Econometric Models and Economic Forecasts, 4th Edition, Irwin/McGraw-Hill
[16] Rosen, D., Saunders, D. (2009). Analytical methods for hedging systematic credit risk with linear factor portfolios. Journal of Economic Dynamics & Control, 33 (2009), 37-52 http://www.r2-financial.com/wp-content/uploads/2010/07/LinearFactor.pdf
[17] Vasicek, O. (2002). Loan portfolio value. RISK, December 2002, 160 - 162.
[18] Wolfinger, R. (2008). Fitting Nonlinear Mixed Models with the New NLMIXED Procedure. SAS Institute Inc. http://www.ats.ucla.edu/stat/sas/library/nlmixedsugi.pdf
[19] Zhang, J., Zhu, F., and Lee, J. (2008). Asset Correlation, Realized Default Correlation, and Portfolio Credit Risk, Moody's KMV, March 2008
http://www.moodysanalytics.com/~/media/Insight/Quantitative- Research/Portfolio-Modeling/08-03-03-Asset-Correlation-and-Portfolio-Risk.ashx

APPENDIX

Proof of Lemma 2.1. Given *s*, we have

$$\Phi(a+bs) = P(\varepsilon \le a+bs)$$
$$= P[(\varepsilon+bs)/\sqrt{1+b^2} \le a/\sqrt{1+b^2}\ ]$$

where $\varepsilon \sim N(0,1)$ is independent of *s*. At this moment, *s* is given as a fixed effect, the only random variable is ε. However, we can view *s* as random when taking expectation $E(\Phi(a+bs))$ with respect to *s*. Since $u = (\varepsilon+bs)/\sqrt{1+b^2}$ is standard normal, $E(\Phi(a+bs))$ is equal to $\Phi(a/\sqrt{1+b^2}\ )$.□

Proof of Proposition 4.1. We have:

$$P(p \le y) = P(z \le \Phi^{-1}(y))$$
$$= P(a+bs \le \Phi^{-1}(y))$$
$$= P(s \le (w-a)/b)$$
$$= \Phi((w-a)/b)$$

where $w = \Phi^{-1}(y)$. The derivative of *w* with respect to *y* is given by $1/\phi(w)$. Thus the derivative of $P(p \le y)$ with respect to *y* is given by

$$\phi((w-a)/b)/(b\phi(w))$$

Replacing *w* by *z*, we have statement (a). For (b), if $\Phi(x)$ is the cumulative distribution for a standard normal variable, then the negative log likelihood reduces to

$$[\Phi^{-1}(p)-a]^2/(2b^2) + \ln(b) - (\Phi^{-1}(p))^2/2$$

This proves (b). □

Proof of Theorem 4.2. By Proposition 4.1 (b), where the parameter b replaces c in Theorem 4.2, the negative log likelihood has the form:

$$\frac{1}{2b^2}\sum_{k=1}^{N} [z_k - (a + \sum_{i=1}^{m} b_i s_{ik})]^2 \quad + N \ln b + C(z) \qquad (8)$$

where $C(z)$ depends only on $\{ z_k \}$. With the maximum likelihood approach, parameters are estimated by minimizing (8). Given b, expression (8) is minimized whenever parameters $a, b_1, b_2, ..., b_m$ (which are independent of b) minimize expression (7). For parameter b, taking the derivative of (8) with respect to b and setting the derivative to zero, we have:

$$b^2 = \frac{1}{N}\sum_{k=1}^{N} [z_k - (a + \sum_{i=1}^{m} b_i s_{ik})]^2$$

This completes the proof. □