Moment Estimators for Autocorrelated Time Series and their Application to Default Correlations

Christoph Frei (cfrei@ualberta.ca)\textsuperscript{1} and Marcus Wunsch (marcus.wunsch@ubs.com)\textsuperscript{2}

\textsuperscript{1}Math. and Stat. Sciences, University of Alberta, Edmonton AB, T6G 2G1, Canada
\textsuperscript{2}UBS AG, Chief Investment Office Wealth Management, 8001 Zurich, Switzerland

August 3, 2017

Abstract

In credit risk modelling, method of moment approaches are popular to estimate latent asset return correlations within and between rating buckets. However, autocorrelation that is often present in default rate time series leads to systematically too low estimations. Adjusting for autocorrelation and shortness of time series, we propose a new estimator. The adjustment is based on convergence and approximation results for general autocorrelated time series. Our adjusted estimator is easily implementable and nonparametric. The adjustment removes a big part of the bias observed in classical estimators due to autocorrelation in default rate time series.

Keywords: autocorrelation, credit risk, latent asset return correlation, method of moments

1 Introduction

In both practice and academia, credit risk of a loan portfolio is modelled frequently based on a structural model generalizing Merton [21] to multiple firms. Obligors in the portfolio are grouped into homogeneous buckets according to their ratings and client segments. It is assumed that each obligor has a latent asset return process that determines the obligor’s default: if the return over the model horizon is below some threshold, the obligor defaults. The return process is driven by systematic and idiosyncratic risk factors which introduce a dependence structure among the obligors’ returns and thus among their defaults. The values of aggregate risk metrics such as Value-at-Risk depend crucially on the latent asset return correlation between obligors of the same bucket (intra-segment correlation) and different buckets (inter-segment correlation). Since the correlation coefficients cannot be measured directly, they are estimated typically via the method of moments or maximum likelihood based on historical default rates within each bucket.

In the classical literature (see Bluhm et al. [3] or McNeil et al. [18] for an overview), these estimators are built on the assumption that the latent asset returns are serially independent. This is in line with the theoretical property that, if all obligors in a bucket have the same
point-in-time rating, all available information on the state of the economy is reflected in the current ratings, and default rates for a given rating bucket will be serially independent. In practice, however, data exhibit autocorrelation in the default rates and thus in the latent asset returns. There are two reasons for this discrepancy. First, in practice, the grouping in rating buckets is often done intentionally by through-the-cycle rather than point-in-time ratings so that default rates reflect the behaviour of economic factors over time, which exhibit autocorrelation. The main motivation for using through-the-cycle parameters is to improve stability of credit risk measures across economic cycles. For instance, the Basel’s advanced internal rating-based approach includes the statement “Although the time horizon used in PD estimation is one year ... banks must use a longer time horizon in assigning ratings”.

A current proposal of the Basel Committee even explicitly asks that assignments to ratings categories generally remain stable over time and through business cycles. Moreover, Moody’s Analytics introduced in 2011 through-the-cycle EDF (Expected Default Frequency) to complement its point-in-time EDF and reduce short-term volatility from the credit cycle. Even when attempting to use point-in-time ratings reflecting the current economic situation, there can still be autocorrelation due to through-the-cycle “dampening” of rating transitions: changes in the credit quality of obligors may not be immediately reflected in their ratings. This affects default rates for the rating buckets, making them dependent on the economic situation and leading to serial dependence in the default rates. The meaning and (dis)advantages of ratings as point-in-time versus through-the-cycle credit indicators are widely discussed in risk management industry and academia; see, for instance, Cantor and Mann [5], Carey and Hrycay [6], Gordy and Howells [11], Heitfield [15] and Löffler [16]. Moreover, the phenomenon of slowly adjusting ratings is well documented and explained, such as in Altman and Rijken [1] and Löffler [17]. Independently of what the reasons of the observed autocorrelation in default rates are, autocorrelation should be taken into account when using default rates as input of estimators.

The goal of this paper is twofold. First, we analyze how autocorrelation affects method-of-moment estimators that are commonly used in industry to determine the latent asset return correlation. While we find that the stationarity and summability of auto-covariances are sufficient for the estimator to converge to the true asset return correlation as the length of the time series goes to infinity, we show that the convergence rate is much slower than in the case of i.i.d. observations. Second, we propose a new estimator that includes correction terms taking the autocorrelation and shortness of the observed time series into account. Our estimator has a simple explicit form, summarized in Appendix A for practical purposes, and we also provide confidence intervals for the estimator. Interestingly, the correction terms

---

1 For example, global annual default rates for the ‘B’ rating category by Standard & Poor’s have lag-1 sample autocorrelation of 46.5% over the years 1981–2016, using data from the S&P 2016 Annual Global Corporate Default Study available at http://www.spglobal.com.


3 “Rating systems should be designed in such a way that assignments to rating categories generally remain stable over time and throughout business cycles. Migration from one category to another should generally be due to idiosyncratic or industry-specific changes rather than due to business cycles.” from Section 4.1 in Basel Committee on Banking Supervision. Consultative document: Reducing variation in credit risk-weighted assets — constraints on the use of internal model approaches, 2016.

4 For more information, please see http://www.moodysanalytics.com/throughthecycleEDF
that our estimator includes are positive when there is positive autocorrelation, showing that asset return correlation and thus typically the values of aggregate risk metrics are underestimated by classical estimators based on the method of moments with serial independence assumptions. The adjustment vanishes as the length of the time series goes to infinity and becomes more dominant the shorter a time series is. For finite time series, the adjustment is present even when asset returns are serially independent, but it is much smaller than in the case of positive autocorrelation. The bias in the estimation of the asset return correlation has been described by Gordy and Heitfield [10] in the case of serially independent asset returns. In this paper, we allow for autocorrelation and, by proposing a new estimator, we provide a remedy for the bias that classical estimators have. We derive the form and properties of our estimator from asymptotic and distributional properties of the sample mean of general autocorrelated time series. Under mild assumptions, we show convergence results for the sample mean in the forms of a law of large numbers and central limit theorem. For autocorrelated time series with finite length, we analyze the difference and bound the error term between (1) expected value of a functional of the sample mean and (2) the functional evaluated at the expected sample mean. This yields an adjustment that is the basis for the correction terms which our new estimator includes.

Maximum likelihood estimators are an alternative to the method of moment that we use. In the case of autocorrelation, however, they lead to highly convoluted integrals, which are computationally not tractable. A sophisticated approximation was proposed by McNeil and Wendin [20]; see also McNeil and Wendin [19] and Wendin [24] for related results and additional details. Their main idea is to apply computational Bayesian estimation (Gibbs sampling) to maximum likelihood estimation of a class of generalized linear mixed models (GLMMs), which are modelling default rates. Our approach has the following advantages. First and foremost, it leads to an easily implementable, explicit formula showing directly the impact of autocorrelation on the value of the estimation. Thus, it is very suitable for industrial applications. Second, our method does not need any distributional properties of the underlying time series and is thus stable, while Gibbs sampling needs assumptions on the prior distribution and the dynamics of the time series (for example, AR(1) process). Finally, our method provides confidence intervals in addition to the point estimates of the latent asset return correlation. This is helpful in practice for sensitivity analyses. On the other hand, the approach by McNeil and Wendin [20] using computational Bayesian estimation allows for a variety of model specifications and priors so that different sources of information and model structures can be used.

The remainder of this paper is organized as follows. We first recall a Merton-based credit risk model in Section 2 and explain why autocorrelation in default rates is an issue. In Section 3, we study properties of the sample mean for a general, autocorrelated time series: we give convergence results in Section 3.1 and analyze in Section 3.2 how these asymptotic results can be adjusted to apply them to finite time series. Section 4 deals with the application of our results to the estimation of the latent asset return correlation in credit risk modelling. We show in Section 4.1 how we obtain new estimators and confidence intervals for the latent asset return correlation, and make in Sections 4.2 and 4.3 comparisons with maximum likelihood estimators for the estimators themselves and their implications on a loss risk metrics. Section 5 concludes and Appendix A summarizes how the formula for our estimator can be applied in practice while Appendix B contains the proof of our main result.
2 Revisiting a structural default model

In this section, we recall method of moments estimators for structural credit risk models of Merton [21] type. Method of moment estimators were introduced in this context by Gordy [9] and Nagpal and Bahar [22] and later refined by Frey and McNeil [8], and Bluhm and Overbeck [2], among others. We refer to Bluhm et al. [3] for an overview of models for correlated defaults, which are typically based on historical default rates as in this paper or proxy variables.

As is widely used in practice and academia, we model credit risk of a portfolio using a structural model generalizing Merton [21] to multiple firms. We group obligors in the portfolio into homogeneous buckets. The normalized asset return of obligor $i$ in bucket $b$ is given by

$$R_i = \sqrt{\varrho_b} Y_b + \sqrt{1 - \varrho_b} \epsilon_i$$

where $\varrho_b \in [0, 1)$, $Y_b$ is a standard normally distributed random variable (the systematic factor of bucket $b$) common to all obligors in bucket $b$ and $\epsilon_i$ is a standard normally distributed random variable (the obligor $i$’s idiosyncratic component) independent of the other $\epsilon_i$ and $Y_b$. For two obligors $i$ and $\tilde{i}$ in different buckets $b$ and $\tilde{b}$, we write

$$\varrho_{b,\tilde{b}} = \text{Corr}(R_i, R_{\tilde{i}}) = \text{Cov}(R_i, R_{\tilde{i}}) = \sqrt{\varrho_b \varrho_{\tilde{b}}} \text{Cov}(Y_b, Y_{\tilde{b}}) = \sqrt{\varrho_b \varrho_{\tilde{b}}} \text{Corr}(Y_b, Y_{\tilde{b}})$$

for the correlation of their latent asset returns. Obligor $i$ defaults if their return is below a threshold $c_b$, which is the same for all obligors in bucket $b$. Hence, if the unconditional default probability of obligors in bucket $b$ is $p_b$, we have

$$p_b = P[R_i \leq c_b] = \Phi(c_b)$$

so that $c_b = \Phi^{-1}(p_b)$. Using the independence of $\epsilon_i$ from $Y_b$, the loss rate in bucket $b$ conditional on the systematic factor $Y_b$ is given by

$$p_b(Y_b) = \Phi\left( \frac{\Phi^{-1}(p_b) - \sqrt{\varrho_b} Y_b}{\sqrt{1 - \varrho_b}} \right). \tag{1}$$

In the following, we focus on the important question of how to estimate $\varrho_b$ and $\varrho_{b,\tilde{b}}$ from historic time series of data. To this end, we consider two buckets $b$ and $\tilde{b}$ with a sufficiently big number of obligors so that the idiosyncratic risk is diversified away on the bucket level. We assume that we are given the default rate times series $(p_b(Y_{b,t}))_{t=1,\ldots,T}$ and $(p_{\tilde{b}}(Y_{\tilde{b},t}))_{t=1,\ldots,T}$ for the two buckets, which correspond to time series $(Y_{b,t})_{t=1,\ldots,T}$ and $(Y_{\tilde{b},t})_{t=1,\ldots,T}$ of systematic factors, reflecting economic conditions. We further assume that $(p_b(Y_{b,t})^2)_{t=1,\ldots,T}$ and $(p_{\tilde{b}}(Y_{\tilde{b},t}) p_b(Y_{b,t}))_{t=1,\ldots,T}$ are stationary, but importantly, they do not need to be serially independent so that they can exhibit autocorrelation, as we observe it in practice.

Proposition 2.5.9 in Bluhm et al. [3] shows that

$$E[(p_b(Y_{b,t}))^2] = \Phi_2(\Phi^{-1}(p_b), \Phi^{-1}(p_{\tilde{b}}); \varrho_b),$$

where $\Phi_2(\ldots; \varrho_b)$ denotes the bivariate normal cumulative distribution function with correlation $\varrho_b$. This implies

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} (p_b(Y_{b,t}))^2 = \Phi_2(\Phi^{-1}(p_b), \Phi^{-1}(p_{\tilde{b}}); \varrho_b) \quad \text{a.s.}$$
by Theorem 3.1 so that we obtain the approximation

\[
\frac{1}{T} \sum_{t=1}^{T} (p_b(Y_{b,t}))^2 \approx \Phi_2\left( \Phi^{-1}\left( \frac{1}{T} \sum_{t=1}^{T} p_b(Y_{b,t}) \right) \right),
\]

where

\[
\Phi^{-1}\left( \frac{1}{T} \sum_{t=1}^{T} p_b(Y_{b,t}) \right) ; \varrho_b.
\]

We consider the function

\[
g_b(\varrho_b) = \Phi_2\left( \Phi^{-1}\left( \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) \right) \right),
\]

for a fixed \( T \) and a realization \( \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) = \frac{1}{T} \sum_{t=1}^{T} p_b(Y_{b,t}(\omega)) \) in some scenario \( \omega \). Restricting \( g_b \) to nonnegative \( \varrho_b \) and fixing a scenario \( \omega \) with \( y_{b,t} = Y_{b,t}(\omega) \), it can be shown that \( g_b \) is invertible\(^5\) so that we obtain an estimator

\[
\hat{\varrho}_{b,1} = g_b^{-1}\left( \frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2 \right).
\]

From Gordy and Heitfield [10], it is known that \( \hat{\varrho}_{b,1} \) is a biased estimator of \( \varrho_b \) because of the finite length of the time series \( (p_b(Y_{b,t}))_{t=1,...,T} \), even when it is serially independent. In contrast, we allow for autocorrelated data, and will see in Section 4 that autocorrelation crucially increases the bias, but there is also a way to correct a large part of the bias.

Similarly to (2) and as in Section 4.1 of Bluhm and Overbeck [2], we can approximate

\[
\frac{1}{T} \sum_{t=1}^{T} p_b(Y_{b,t})p_b(Y_{\tilde{b},t}) \approx \Phi_2\left( \Phi^{-1}\left( \frac{1}{T} \sum_{t=1}^{T} p_b(Y_{b,t}) \right) \right),
\]

and thus obtain for \( \varrho_{b,\tilde{b}} \), an estimator

\[
\hat{\varrho}_{b,\tilde{b},1} = g_{b,\tilde{b}}^{-1}\left( \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t})p_b(y_{\tilde{b},t}) \right),
\]

where \( g_{b,\tilde{b}}^{-1} \) is the inverse function of

\[
g_{b,\tilde{b}}(\varrho_{b,\tilde{b}}) = \Phi_2\left( \Phi^{-1}\left( \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) \right) \right),
\]

restricted on nonnegative \( \varrho_{b,\tilde{b}} \), with \( y_{b,t} \) and \( y_{\tilde{b},t} \) realizations of \( Y_{b,t} \) and \( Y_{\tilde{b},t} \), respectively, in some scenario \( \omega \).

\(^5\) \( g_b \) is strictly increasing as it will follow later from (11) and \( g_b(0) = \left( \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) \right)^2 \leq \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t})^2 \) by Jensen’s inequality and \( g_b(1) = \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) \geq \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t})^2 \) since \( p_b(y_{b,t}) \in [0, 1] \).
3 Convergence results for autocorrelated time series

In this section, we analyze the behavior of the sample mean \( \frac{1}{T} \sum_{t=1}^{T} Z_t \) of an autocorrelated time series \((Z_t)_{t=-\infty}^{\infty}\). We first show in Section 3.1 convergence results for the limit of the sample mean as \( T \to \infty \) under weak assumptions and then derive in Section 3.2 results for finite \( T \). The results of this section can be generalized to multidimensional processes, i.e., when \( Z_t \) is multidimensional where the components of \( Z_t \) can be correlated. However, since this leads to cumbersome notation and does not give new insights, we restrict the presentation to a one-dimensional \( Z_t \).

3.1 Asymptotic properties

We impose the following assumptions:

A1 Stationarity: any \( k \) subsequent random variables \( Z_{t+1}, \ldots, Z_{t+k} \) have the same distribution regardless of the starting point \( t \).

A2 Absolute summability of autocovariances: there exists a constant \( c < \infty \) such that

\[
\sum_{t=-\infty}^{\infty} |\gamma_{s,t}| \leq c \quad \text{for all } s,
\]

where the autocovariances \( \gamma_{s,t} \) are defined by

\[
\gamma_{s,t} = E\left[ (Z_s - E[Z_s]) (Z_t - E[Z_t]) \right].
\]

The next result follows from an application of the ergodic theorem; see, for example, Theorem 19.1 in Greene [13].

**Proposition 3.1.** Under assumptions A1 and A2, the sample mean of the time series \((Z_t)_{t=-\infty}^{\infty}\) converges almost surely:

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} Z_t = E[Z_1]
\]

We can invoke this result also for the convergence of the sample variances. For this case to hold true, we presume that all of the three versions of the auto-cokurtosis of \( Y \) are summable, as we explain in the next example.

As an example, consider the AR(1) process

\[
X_t = \rho X_{t-1} + (1 - \rho) \mu + \sigma \epsilon_t
\]

for \(|\rho| < 1\), \( \mu \in \mathbb{R} \), \( \sigma > 0 \), and independent, standard normally distributed \( \epsilon_t \). This process can be demeaned by subtracting the mean \( \mu \), i.e., the process \( Z_t = X_t - \mu \) has zero mean.

\[\text{In this case, the AR(1) process is called wide-sense stationary.}\]
Note that for each \( t \), \( Z_t \) is normally distributed with zero mean and variance \( \frac{\sigma^2}{1 - \rho^2} \), so that \( Z_t \) is identically distributed, but not independent, as the autocovariance satisfies

\[
\gamma_{s,t} = \frac{\sigma^2}{1 - \rho^2} \rho^{|s-t|}.
\]

Since thus assumptions A1 and A2 are satisfied, the strong law of large numbers from Proposition 3.1 holds for AR(1) process.

Because \( Z_s \) and \( Z_t \) are normally distributed, all comoments of \( Z_s \) and \( Z_t \) are finite. In particular, if we define an autocokurtosis \( \kappa_{s,t} = \kappa_{s,t}^{(2,2)} \) as

\[
\kappa_{s,t} = \frac{E[(Z_s)^2(Z_t)^2]}{\gamma_{s,s} \gamma_{t,t}}
\]

then

\[
\kappa_{s,t} = \frac{3\sigma^4}{1 - \rho^4} \frac{\rho^{(2|s-t|)}}{\sigma^4(1 - \rho^2)^2} = \frac{3(1 - \rho^2)^2}{1 + \rho^2} \rho^{(2|s-t|)}
\]

for integers \( s \) and \( t \). Similar (summable) expressions hold for the auto-cokurtoses \( \kappa_{s,t}^{(1,3)} \) and \( \kappa_{s,t}^{(3,1)} \). This implies that we can invoke Proposition 3.1 for the almost sure convergence of the second sample moment of an AR(1) process as well.

More generally, almost sure convergence of all sample moments of an ARMA(\( p, q \)) process holds if the stationarity condition (zeros of its AR(\( p \)) polynomial lie outside the unit circle) is satisfied; compare Sections 3.A and 7.2 in Hamilton [14].

For the later application, we mention also a central limit theorem that can be used in our situation. In addition to the assumptions of stationarity and absolute summability of the autocovariances, we assume that

A3 Asymptotic uncorrelatedness: \( E[Z_t|Z_{t-k}, Z_{t-k-1}, \ldots] \) converges in mean square to zero as \( k \to \infty \).

A4 Asymptotic negligibility of innovations: \( \sum_{k=0}^{\infty} E[r_{t,k}^2] \) is finite for fixed \( t \), where

\[
r_{t,k} = E[Z_t|Z_{t-k}, Z_{t-k-1}, \ldots] - E[Z_t|Z_{t-k-1}, Z_{t-k-2}, \ldots].
\]

An AR(1) process

\[
Z_t = \rho Z_{t-1} + \epsilon_t
\]

with mean zero and \( |\rho| < 1 \) satisfies assumptions A3 and A4. Indeed, we can use Wold’s representation that

\[
Z_t = \sum_{j=0}^{\infty} \rho^j \epsilon_{t-j},
\]

Note that in addition to this version of the autocokurtosis with powers (2, 2), there are two other versions with powers (1, 3) and (3, 1).
where the convergence of the series is in mean square (and almost surely). It now follows that
\[ E[Z_t|Z_{t-k}, Z_{t-k-1}, \ldots] = \sum_{j=k}^{\infty} \rho^j \epsilon_{t-j} \]
indeed converges to zero in mean square as the remainder of a converging series. Because
\[ E[Z_t|Z_{t-k}, Z_{t-k-1}, \ldots] - E[Z_t|Z_{t-k-1}, Z_{t-k-2}, \ldots] = \rho^k \epsilon_{t-k}, \]
A4 is satisfied with \( r_{t,k} = \rho^k \epsilon_{t-k} \) since
\[ \sum_{k=0}^{\infty} E[r_{t,k}^2] = \sum_{k=0}^{\infty} \rho^{2k} E[\epsilon_{t-k}^2] = \frac{1}{1 - \rho^2} < \infty. \]

3.2 Adjusting for shortness and autocorrelation

We now restrict ourselves to the case where we have \( T < \infty \) observations of the time series. We consider a moment estimator of the form \( g(\theta) = \mu \), where \( \theta \) is a parameter to be estimated, \( \mu \) is the unknown mean of the stationary time series \( (Z_t)_{t=1, \ldots, T} \). We assume that \( g \) is three times continuously differentiable and invertible with inverse \( \tilde{g} = g^{-1} \).

A natural estimator for \( \theta \) is the moment estimator
\[ \hat{\theta}_1 = \tilde{g}\left( \frac{1}{T} \sum_{t=1}^{T} Z_t \right). \]

By continuity of \( \tilde{g} \) and the law of large numbers (compare Theorem 3.1), \( \hat{\theta}_1 \) converges almost surely to \( \theta \) as \( T \to \infty \). However, for finite \( T \), there is an estimation bias because
\[ E[\hat{\theta}_1] = E \left[ \tilde{g}\left( \frac{1}{T} \sum_{t=1}^{T} Z_t \right) \right] \neq \tilde{g} \left( E \left[ \frac{1}{T} \sum_{t=1}^{T} Z_t \right] \right) = \tilde{g}(\mu) = \theta. \]

Equality would hold only if \( \tilde{g} \) were linear. The next theorem, whose proof is in Appendix B, allows us to improve the estimation.
Theorem 3.3. Under assumption A1, there exists a random variable $\xi$ with values between $\mu$ and $\frac{1}{T} \sum_{t=1}^{T} Z_t$ such that

$$
\left| \theta - E\left[ \tilde{g}\left( \frac{1}{T} \sum_{t=1}^{T} Z_t \right) \right] - \frac{g''(\theta)}{2T(g'(\theta))^3} \text{Var}(Z_1) - \frac{g''(\theta)}{T^2(g'(\theta))^3} \sum_{\ell=1}^{T-1} (T - \ell) \text{Cov}(Z_1, Z_{1+\ell}) \right|
$$

$$
\leq \frac{1}{6} E[ (\tilde{g}''(\xi))^4 ]^{1/4} E \left[ \left( \frac{1}{T} \sum_{t=1}^{T} Z_t - \mu \right)^4 \right]^{3/4}.
$$

The theorem says that $\theta$ can be approximated by

$$
E \left[ \tilde{g}\left( \frac{1}{T} \sum_{t=1}^{T} Z_t \right) \right] + \frac{g''(\theta)}{2T(g'(\theta))^3} \text{Var}(Z_1) + \frac{g''(\theta)}{T^2(g'(\theta))^3} \sum_{\ell=1}^{T-1} (T - \ell) \text{Cov}(Z_1, Z_{1+\ell}),
$$

and there is an explicit upper bound for the resulting estimation error. To make use of this result, we choose a smaller number $k$ of terms in the sum in (7) and make the following replacements: $\theta$ is replaced by the estimator $\hat{\theta}_1 = \tilde{g}\left( \frac{1}{T} \sum_{t=1}^{T} Z_t \right)$, and $\text{Var}(Z_1)$ and $\text{Cov}(Z_1, Z_{1+\ell})$ are replaced by sample variance and sample autocovariances. This gives us a new estimator

$$
\hat{\theta}_2 = \tilde{g}(\bar{\mu}) + \frac{g''(\tilde{g}(\bar{\mu}))}{2T(g'(\tilde{g}(\bar{\mu})))^3} \alpha_0 + \frac{g''(\tilde{g}(\bar{\mu}))}{T(g'(\tilde{g}(\bar{\mu})))^3} \sum_{\ell=1}^{k} (1 - \ell/T) \alpha_\ell,
$$

where $\bar{\mu} = \frac{1}{T} \sum_{t=1}^{T} Z_t$ is the sample mean and

$$
\alpha_\ell = \frac{1}{T} \sum_{t=1+\ell}^{T} (Z_t - \bar{\mu})(Z_{t-\ell} - \bar{\mu}), \quad \ell = 0, 1, \ldots, k
$$

is the lag-$\ell$ sample autocovariance. As usual in time series analysis, we divide in the definition of the sample autocovariance by $T$ and not $T - \ell$; compare, for instance, Section 1.6 in Shumway and Stoffer [23].

In addition to the bounds for the point estimates, we can also find approximate confidence intervals. For this, we note that

$$
\text{Var}\left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t \right) = \frac{1}{T} \text{Var} \left( \sum_{t=1}^{T} Z_t \right) = \frac{1}{T} \sum_{i,j=1}^{T} \text{Cov}(Z_i, Z_j)
$$

$$
= \frac{1}{T} \sum_{t=1}^{T} \text{Var}(Z_t) + \frac{2}{T} \sum_{\ell=1}^{T-1} \sum_{t=1}^{T-\ell} \text{Cov}(Z_t, Z_{t+\ell})
$$

$$
= \text{Var}(Z_1) + \frac{2}{T} \sum_{\ell=1}^{T-1} (T - \ell) \text{Cov}(Z_1, Z_{1+\ell})
$$

$$
\approx \alpha_0 + \frac{2}{T} \sum_{\ell=1}^{k} (T - \ell) \alpha_\ell.
$$
Applying Gordin’s central limit theorem (5) and using the delta method, we can approximate a \( p \% \) confidence interval for \( \hat{\theta}_2 \) by

\[
\hat{\theta}_2 \pm \frac{q|\hat{g}'(\hat{\mu})|}{\sqrt{T}} \sqrt{\alpha_0 + 2 \sum_{\ell=1}^{k} (1 - \ell/T)\alpha_{\ell}}
\]

where \( q = t_{T-1}^{-1}(1 - (1 - p/100)/2) \) with \( t_{T-1}^{-1} \) denoting the inverse of the cumulative distribution function of the \( t \)-distribution with \( T - 1 \) degrees of freedom. The approximate confidence interval (10) can be easily implemented and is suitable for our purposes. An alternative could be derived by using blockwise bootstrap as in Götze and Künsch [12], who show that their methods lead to gains in asymptotic accuracy.

**Remark 3.4.** In (8) and (10), one could use the idea of tapering by replacing the terms \( 1 - \ell/T \) by weights \( w_\ell \) which decay smoothly from 1 for \( \ell = 0 \) to 0 for \( \ell = k + 1 \). This would reduce the jump in the weight \( 1 - k/T \) for \( \alpha_k \) to the weight 0 for \( \alpha_{k+1} \) in the sums in (8) and (10). For a statistical method on how to select the number \( k \) of terms in the sums in (8) and (10), we refer to Bühlmann [4], who proposes an iterative plug-in scheme for the locally optimal window width in nonparametric estimations.

### 4 Application to credit risk

We show in Section 4.1 how the accuracy of the method of moments estimators from Section 2 can be improved by using the results of Section 3.2. In Section 4.2, we give a performance comparison of the original and adjusted method of moment estimators with a maximum likelihood estimator (MLE). Section 4.3 gives an example how the estimated parameter can be used in the computation of a loss risk metrics and compares the metrics for the different estimators.

#### 4.1 New estimators for the latent asset return correlation

Let us consider the setting of Section 2 with two rating buckets \( b \) and \( \tilde{b} \) that have a sufficiently big number of obligors so that the idiosyncratic risk is diversified away on the bucket level. To apply the results of Section 3.2, we note that the estimators (3) and (4) are of the form

\[
\hat{\theta}_1 = \tilde{g} \left( \frac{1}{T} \sum_{t=1}^{T} Z_t(\omega) \right)
\]

as realizations of (6) with \( \tilde{g} = g_{b,1}^{(-1)} \), \( \hat{\theta}_1 = \hat{\theta}_{b,1} \), \( Z_t = (p_b(Y_{b,t}))^2 \) and \( \tilde{g} = g_{b,\tilde{b},1}^{(-1)} \), \( \hat{\theta}_1 = \hat{\theta}_{b,\tilde{b},1} \), \( Z_t = p_{b}(Y_{b,t})p_{\tilde{b}}(Y_{b,t}) \), respectively. Since, in either case, the inverse of \( \tilde{g} \) is \( g(\theta_1) = \Phi_2(s, t; \theta_1) \), we calculate the derivatives

\[
\frac{\partial}{\partial \theta_1} \Phi_2(s, t; \theta_1) = \phi_2(s, t; \theta_1) = \frac{1}{2\pi \sqrt{1 - \theta_1^2}} \exp \left( -\frac{s^2/2 - \theta_1 st + t^2/2}{1 - \theta_1^2} \right),
\]

\[
\frac{\partial^2}{(\partial \theta_1)^2} \Phi_2(s, t; \theta_1) = \frac{st + \theta_1(1 - s^2 - t^2) + st\theta_1^2 - \theta_1^3}{2\pi(1 - \theta_1^2)^{5/2}} \exp \left( -\frac{s^2/2 - \theta_1 st + t^2/2}{1 - \theta_1^2} \right).
\]
From Section 3.2, we can find estimated correlations \( \hat{\rho}_{b,1}^2 \) and \( \hat{\rho}_{b,\tilde{b},1}^2 \); taking the short length and autocorrelation of the time series into account, by

\[
\hat{\rho}_{b,1} = \hat{\rho}_{b,1} + \frac{g''(\hat{\rho}_{b,1})}{T(g'(\hat{\rho}_{b,1}))^3} \left( \alpha_{b,0}/2 + \sum_{\ell=1}^{k} (1 - \ell/T)\alpha_{b,\ell} \right),
\]

\[
\hat{\rho}_{b,\tilde{b},1} = \hat{\rho}_{b,\tilde{b},1} + \frac{g''(\hat{\rho}_{b,\tilde{b},1})}{T(g'(\hat{\rho}_{b,\tilde{b},1}))^3} \left( \alpha_{b,\tilde{b},0}/2 + \sum_{\ell=1}^{k} (1 - \ell/T)\alpha_{b,\tilde{b},\ell} \right),
\]

where

- \( \hat{\rho}_{b,1} \) and \( \hat{\rho}_{b,\tilde{b},1} \) are the estimations from (3) and (4);
- \( g'(\hat{\rho}_{b,1}) \) and \( g''(\hat{\rho}_{b,1}) \) equal to (11) and (12) with \( \rho_1 = \hat{\rho}_{b,1} \) and evaluated at \( s = t = \Phi^{-1}(\frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t})) \); analogously, \( g'(\hat{\rho}_{b,\tilde{b},1}) \) and \( g''(\hat{\rho}_{b,\tilde{b},1}) \) equal to (11) and (12) with \( \rho_1 = \hat{\rho}_{b,\tilde{b},1} \) and evaluated at \( s = \Phi^{-1}(\frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t})) \) and \( t = \Phi^{-1}(\frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t})) \); 
- \( \alpha_{b,\ell} \) and \( \alpha_{b,\tilde{b},\ell} \) are the lag-\( \ell \) sample autocovariances of the time series \((p_b(y_{b,t}))_{t=1,\ldots,T}\) and \((p_b(y_{b,t})p_b(y_{\tilde{b},t}))_{t=1,\ldots,T}\), respectively, i.e.,

\[
\alpha_{b,\ell} = \frac{1}{T} \sum_{t=1+\ell}^{T} \left( (p_b(y_{b,t}))^2 - \mu_b \right) \left( (p_b(y_{b,t-\ell}))^2 - \mu_b \right),
\]

\[
\alpha_{b,\tilde{b},\ell} = \frac{1}{T} \sum_{t=1+\ell}^{T} \left( p_b(y_{b,t})p_b(y_{\tilde{b},t}) - \mu_{b,\tilde{b}} \right) \left( (p_b(y_{b,t-\ell})p_b(y_{\tilde{b},t-\ell})) - \mu_{b,\tilde{b}} \right),
\]

where \( \mu_b = \frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2 \) and \( \mu_{b,\tilde{b}} = \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t})p_b(y_{\tilde{b},t}). \)

Figures 1 and 2 illustrate how the adjustment affects the estimations in the cases where the underlying factor is modelled by independent observations or an AR(1) process. For the AR(1) process, the bias of the original estimator \( \hat{\rho}_{b,1} \) is much bigger than in the case of independent observations. Therefore, our proposed adjustments yield a particularly big improvement for autocorrelated time series, but still improve the classical estimator in the case of independent observations. We display here plots for the correlation estimation within a bucket and an underlying true correlation coefficient of \( \rho_b = 0.05 \), which is typical for a large, well diversified credit portfolio, but we made similar observations when considering the correlation between two buckets or when choosing different correlation coefficients.

To improve the estimator \( \hat{\rho}_{b,1} \), we focused on the approximation error coming from \( \frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2 \) which appears to be crucial while there is another approximation error from replacing \( p_b \) by \( \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) \) in (2), which however has a smaller impact on the estimation \( \hat{\rho}_{b,1} \) because of its appearance as argument in \( \Phi_2 \) compared to the more sensitive dependence of \( g^{-1} \) on \( \frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2 \). A multidimensional generalization of our results would build an estimator that simultaneously improves the approximation error from both \( \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) \) and \( \frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2 \). We restrain from using such an estimator because we observed that the additionally gained precision is small for the estimation, the notation is
Figure 1: Comparison of distributions of the original correlation estimation and the adjusted correlation estimation of different orders $k = 1, \ldots, 5$. The underlying factors are simulated based on an AR(1) process with coefficient 0.7 (upper panel) and i.i.d. observations (lower panel). The solid curves are density estimations and the dotted lines give sample means. The adjustments remove a big part of the bias so that the adjusted means are much closer to the true value of $\rho_b = 0.05$, for both the AR(1) process and the i.i.d. observations. The time series have length 80 (reflecting 20 years of data, assuming that each time step corresponds to a quarter), the quarterly probability of default is set equal to 0.2%, and the plots are based on 50,000 simulations for each AR(1) and i.i.d. time series.
Figure 2: Comparison of the original correlation estimation and the adjusted correlation estimation of orders $k = 5$ for different lengths of the time series. The underlying factors are simulated based on an AR(1) process with coefficient 0.7 (upper panel) and i.i.d. observations (lower panel). The solid curves are correlation estimations. Again, the adjustment removes a big part of the bias so that the adjusted mean is much closer to the true value of $\rho_b = 0.05$, for both the AR(1) process and the i.i.d. observations. An additional advantage of our method is that it allows us to compute an approximate 95% confidence intervals, showed by the dashed lines. The quarterly probability of default is set equal to 0.2%, and the plots are based on 1,000 simulations for each AR(1) and i.i.d. time series.
and Footnote 5, there exists a function $h$ that is large extent in the first moment. Assuming that the idiosyncratic risk for different periods is independent, it will average out to $\frac{q}{\sqrt{T}} \frac{\bar{d}(\hat{\mu})}{\sqrt{T}} \sqrt{\alpha_0 + 2 \sum_{t=1}^{k} (1 - \ell/T) \alpha_t}$ in (10). From

$$\frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2 = \Phi_2 \left( \Phi^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) \right) ; \Phi^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) \right) ; \hat{\rho}_b \right)$$

and Footnote 5, there exists a function $h$ with $h \left( \frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2, \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t}) \right) = \hat{\rho}_b$. While $h$ is not explicitly given, we can numerically calculate its partial derivatives $\frac{\partial h}{\partial x}$ and $\frac{\partial h}{\partial y}$.

For the approximate confidence intervals in Figure 2, we set $s_b$ equal to the nonnegative square root of

$$s_b^2 = \frac{q^2}{T} \left( \frac{\partial h}{\partial x} \right)^2 \left( \alpha_{b,0} + 2 \sum_{t=1}^{k} (1 - \ell/T) \alpha_{b,t} \right) + \left( \frac{\partial h}{\partial y} \right)^2 \left( \beta_{b,0} + 2 \sum_{t=1}^{k} (1 - \ell/T) \beta_{b,t} \right) + 2 \frac{\partial h}{\partial x} \frac{\partial h}{\partial y} \left( \gamma_{b,0} + 2 \sum_{t=1}^{k} (1 - \ell/T) \gamma_{b,t} \right)$$

where

$$\beta_{b,t} = \frac{1}{T - 1 - \ell} \sum_{t=1+\ell}^{T} (p_b(y_{b,t}) - \bar{\rho}_b) (p_b(y_{b,t-\ell}) - \bar{\rho}_b)$$

$$\gamma_{b,t} = \frac{1}{T - 1 - \ell} \sum_{t=1+\ell}^{T} (p_b(y_{b,t}) - \bar{\rho}_b) (p_b(y_{b,t-\ell}) \bar{\rho}_b)$$

with $\bar{\rho}_b = \frac{1}{T} \sum_{t=1}^{T} p_b(y_{b,t})$ are defined similarly to (13). Taking changes in both arguments of $h$ into consideration helps us get more precise approximate confidence intervals in Figure 2.

As mentioned earlier, the simulations in this section are done for a bucket with a sufficiently large number of obligors. When there is a smaller number of obligors, the estimators can still be applied, but lead to an additional error because of the idiosyncratic risk. We next analyze this error and discuss how it can be partially corrected. When applying the estimators to the latent asset return correlation in a bucket in the presence of idiosyncratic risk, the estimated values will typically be conservative, i.e., too high on average compared to the limiting case of infinitely many obligors. The intuition behind this is as follows. Let $x_1, \ldots, x_T$ be a default rate time series, reflecting both idiosyncratic and systematic risks. Recall from (3) that the moment estimator $\hat{\rho}_b$ is defined by

$$\Phi_2 \left( \Phi^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} x_t \right) \right) = \frac{1}{T} \sum_{t=1}^{T} x_t^2.$$  \hspace{1cm} (14)

Assuming that the idiosyncratic risk for different periods is independent, it will average out to a large extent in the first moment $\frac{1}{T} \sum_{t=1}^{T} x_t$, particularly, for bigger values of $T$. Therefore,
\[ \frac{1}{T} \sum_{t=1}^{T} x_t \] is similar to its analogue in a model with infinitely many obligors. In contrast, the idiosyncratic risk contributes to the second moment \[ \frac{1}{T} \sum_{t=1}^{T} x_t^2 \]. Compared to the case without idiosyncratic risk, \[ \frac{1}{T} \sum_{t=1}^{T} x_t \] on the left-hand side of (14) is essentially unchanged while the right-hand side of (14) is increased on average. Since the function \( \Phi_2 \) is increasing in \( \rho_b \) (compare (11)), this will typically lead to a bigger estimation \( \hat{\rho}_b \). This additional bias resulting from the finite number \( N \) of obligors can be partially corrected. Indeed, writing \( X_t = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\sqrt{\rho_b} Y_{i,t} + \sqrt{1-\rho_b} \epsilon_{i,t} \leq \Phi^{-1}(p_b)} \) in the notation of Section 2, the squared default rate \( X_t^2 \) conditional on the systematic factor \( Y_{b,t} \) can be computed as

\[ E[X_t^2|Y_{b,t}] = \left( 1 - \frac{1}{N} \right) (p_b(Y_{b,t}))^2 + p_b(Y_{b,t}) \frac{1}{N} \text{ for } p_b(Y_{b,t}) = \Phi \left( \frac{\Phi^{-1}(p_b) - \sqrt{\rho_b} Y_{b,t}}{\sqrt{1-\rho_b}} \right) , \]

hence the squared default rate for infinitely many obligors satisfies

\[ (p_b(Y_{b,t}))^2 = \frac{1}{1 - 1/N} E[X_t^2|Y_{b,t}] - \frac{1/N}{1 - 1/N} p_b(Y_{b,t}) \approx E[X_t^2|Y_{b,t}] - \frac{1}{N} E[X_t|Y_{b,t}] . \quad (15) \]

Therefore, to take the influence of a small number of obligors into account, we can replace \( \frac{1}{T} \sum_{t=1}^{T} x_t^2 \) in (14) by \( \frac{1}{T} \sum_{t=1}^{T} (x_t^2 - x_t/N) \). The plots in Figure 3 show that this additional adjustment works well: without the adjustment the latent asset return correlation is overestimated when the number of obligors is not sufficiently big. The additional adjustment brings the estimates closer to those in the limiting case of an infinite number of obligors. Note that the argumentation of a typically higher estimate for a finite number of obligors and its adjustment need a sufficiently big \( T \) (80 in Figure 3) and apply only to the latent asset return correlation within a bucket, but not for that between two buckets. For the correlation estimation between two buckets, \( \frac{1}{T} \sum_{t=1}^{T} x_t^2 \) in (14) is replaced by \( \frac{1}{T} \sum_{t=1}^{T} x_t \tilde{x}_t \) where \( \tilde{x}_1, \ldots, \tilde{x}_T \) is the default rate time series of the other bucket. Because a large part of the idiosyncratic risk in \( \frac{1}{T} \sum_{t=1}^{T} x_t \tilde{x}_t \) averages out thanks to its independence across buckets and time periods, the estimation for the latent asset return correlation between two buckets is much less affected by idiosyncratic risk than that within a bucket.

**Remark 4.1.** We used the classical setting that obligors are grouped into homogeneous buckets according to their ratings and client segments, and modelled the asset return correlation of each bucket independently. In applications, it can be useful to impose that the buckets in a client segment across different rating classes have the same correlation. We briefly mention how this can be addressed in our approach. Since in this case one cannot find a correlation parameter individually for each bucket, a natural approach is to introduce weights \( w_b \geq 0 \) with \( \sum_{b \in B} w_b = 1 \) for the different buckets according to their importance (for example, according to the total exposure at default per bucket) and to minimize over \( \rho_1 \) the sum of squared errors

\[ \sum_{b \in B} w_b \left[ \rho_1 - g_b^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2 \right) \right]^2 . \]

The optimal \( \rho_1 \) is given by the first-order condition \( \sum_{b \in B} w_b \left[ \rho_1 - g_b^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2 \right) \right] = 0 \), hence \( \rho_1 = \sum_{b \in B} w_b g_b^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} (p_b(y_{b,t}))^2 \right) \). Because this is of the same form as (3), just with a linear combination \( \sum_{b \in B} w_b g_b^{-1} \) instead of a single \( g_b^{-1} \), we can argue similarly as above to obtain estimators taking the short length and autocorrelation of the time series into account. Since this does not offer new insights, we do not spell out the details here.
Figure 3: Influence of the idiosyncratic risk on the correlation estimations, as a function of the number of obligors in the bucket. For a finite number of obligors, idiosyncratic risk is present in the default rate time series, leading typically to higher correlation estimates (dashed curves, which are based on (14)) than for an infinite number of obligors. The solid curves are additionally corrected for finite bucket size, as discussed after (15). They are stable regarding the number of obligors, and the means of the adjusted estimators are close to the true value $\varrho_b = 0.05$. As in Figure 1, the underlying factors are simulated based on an AR(1) process with coefficient 0.7 (upper panel) and i.i.d. observations (lower panel), the time series have length 80, and the quarterly probability of default is 0.2%. The plots are based on 5,000 simulations for each AR(1) and i.i.d. time series. The chosen low values for quarterly default probability (0.2%) and $\varrho_b = 0.05$ mean that the idiosyncratic risk has a big influence on the correlation estimation when the number of obligors is not sufficiently big.
4.2 A performance comparison with an MLE

MLE is an alternative to method of moment estimators, but it cannot be easily adjusted for autocorrelation in the underlying data. If we consider a single bucket, an MLE for the pairwise asset return correlations consists in finding the argument $\varrho_b^* \in [0, 1)$ that maximizes the likelihood function

$$L(\varrho_b) = \int_{\mathbb{R}^T} p_b(y_t)^{D_t} \left(1 - p_b(y_t)\right)^{N_t-D_t} dF(y_1, \ldots, y_T)$$

(16)

where $D_t$ is the number of obligor defaults at time $t$, $N_t$ is the total number of obligors at time $t$, $p_b(y_t)$ is the conditional loss rate given in (1), and $F$ is the cumulative distribution function of the joint distribution of $(Y_1, \ldots, Y_T)$.

Since (16) requires us to specify the joint distribution function $F$, the MLE fundamentally rests upon distributional assumptions on the systematic factor $(Y_1, \ldots, Y_T)$. Let us next consider two examples.

**Example I: Independent Gaussian systematic returns.** If we assume that the latent asset returns $(Y_t)_{t=1,\ldots,T}$ are i.i.d. Gaussian, then the likelihood function (16) simplifies to

$$L(\varrho_b) = \prod_{t=1}^T \left[ \int_{-\infty}^{\infty} p(y_t)^{D_t} \left(1 - p(y_t)\right)^{N_t-D_t} d\Phi(y_t) \right],$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution.

**Example II: AR(1) systematic returns.** On the other hand, under the assumption that $(Y_t)_{t=1,\ldots,T}$ is governed by AR(1) dynamics, i.e.,

$$X_t^{(k)} = \sqrt{\varrho} Y_t + \sqrt{1-\varrho} \Xi_t^{(k)}, \quad k = 1, \ldots, N_t,$$

$$Y_t = \rho Y_{t-1} + \sqrt{1-\rho^2} \Upsilon_t, \quad Y_{t-1} \perp \Upsilon_t, \quad t = 2, \ldots, T,$$

the likelihood function takes the form

$$L(\varrho, \rho, \mu) = \int_{\mathbb{R}^T} p(y_t)^{D_t} \left(1 - p(y_t)\right)^{N_t-D_t} d\Phi_T((y_1, \ldots, y_T)'; \mu, \Sigma_T),$$

(17)

where $\Phi_T(\cdot; \mu, \Sigma)$ is the cumulative distribution function of the $T$-dimensional Gaussian distribution with mean $\mu$ and covariance matrix $\Sigma$. Observe that $\Sigma_T$ in (17) is given by

$$\Sigma_T = \begin{pmatrix} 1 & \rho & \cdots & \rho^T \\ \rho & 1 & \cdots & \rho^{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^T & \rho^{T-1} & \cdots & 1 \end{pmatrix}$$

as $(Y_1, \ldots, Y_T) \sim \mathcal{N}(0, \Sigma_T)$.

In general, the MLE for the estimation of latent asset returns involves as likelihood function a $T$-dimensional integral. In the case of an AR(1) process, the conditional likelihood
at a time $t$ given information observable at $t - 1$ results in a one-dimensional integral when the idiosyncratic risk is diversified away because of a big number of obligors in the rating bucket. Therefore, the $T$-dimensional integral in (17) can be written as an iteration of $T$ one-dimensional integrals. Note, however, that this needs the knowledge or assumption that the systematic returns follow an AR(1) process. In general, MLE for dependent processes is computationally very demanding and can only be approximated by using sophisticated numerical techniques (see McNeil and Wendin [20], and Wendin [24]). Moreover, these methods require assumptions on the prior distribution and the dynamics of the time series, which typically is unknown in practice. In contrast, our adjusted estimator is nonparametric in the sense that it does not require knowledge of the dynamics of the underlying time series.

We conclude this section with Figure 4, which gives a performance comparison of method of moment (MoM) and MLE for an AR(1) process and i.i.d. observations. In both cases, we choose the MLE (16), assuming serial independence since we suppose that the underlying autocorrelation is not known. For i.i.d. observations, we find that all three estimators (original and adjusted MoM estimators, and MLE) perform well and similarly. Indeed, the MLE has slightly smaller standard deviation than the MoM estimators. The bias of MLE is between that of the original and adjusted MoM estimators, but in all three cases, the bias is small (less than 3.5%) for i.i.d. observations. For the AR(1) process, however, the adjusted MoM estimator has a much smaller bias (2.2%) than the original MoM estimator (bias: 11.5%) and the MLE (bias: 9.2%). In the upper panel of Figure 4, we see particularly that the MLE has a higher risk of extremely underestimating the true correlation value, compared to the MoM estimators. Standard deviations across the three estimators are similar, with original MoM and MLE on the same level, and the adjusted MoM slightly higher. In summary, our adjusted estimator is easy to implement and shows a clear performance improvement in terms of reduced bias, compared to both original MoM estimator and MLE when the dynamics of the underlying time series is not explicitly known.

4.3 Application to loss risk metrics

Estimators for the latent asset return correlation are often used to compute risk metrics on loan portfolios. In this section, we first recall how this can be done and then analyze in an example how the choice of the estimator affects the values of a risk metrics.

To compute a risk metrics, we position ourselves in the situation where we are given a default time series over the last 80 quarters. As is often used in practice and in line with the Basel II capital accord for banking-book transactions\footnote{Paragraph 178 in Basel Committee on Banking Supervision. International convergence of capital measures and capital standards, 2006.\)}, we compute the Value-at-Risk (VaR) over a forward-looking one-year period. To this end, we estimate the latent asset return correlation using one of the previously discussed estimators as well as the autocorrelation based on data over the last 80 quarters. The autocorrelation is estimated in the standard way, after transforming the default time series to realizations of the systematic factor by applying the inverse of (1). We then use the estimated parameters for a Monte Carlo simulation with 50,000 paths of the default time series over the next four quarters. For this example, we consider a bucket with a total loan volume of $1 billion with a sufficiently large number of
Figure 4: Comparison of distributions of the original method of moment (MoM) estimator, the 5th-order adjusted MoM estimator and an MLE. The underlying factors are simulated based on an AR(1) process with coefficient 0.7 (upper panel) and i.i.d. observations (lower panel). The solid curves are density estimations and the dotted lines give sample means. In the case of i.i.d. observations, the three estimators perform similarly: original MoM (bias = 3.5%, std. dev. = 0.0111), adjusted MoM (bias = 1.2%, std. dev. = 0.0114), and MLE (bias = 2.6%, std. dev. = 0.0085). For the AR(1) process, the bias of the MLE (bias = 9.2%, std. dev. = 0.0135) is similar to that of the original MoM estimator (bias = 11.5%, std. dev. = 0.0135), and much higher than that of the adjusted MoM estimator (bias = 2.2%, std. dev. = 0.0165). The time series have length 80 (reflecting 20 years of data, assuming that each time step corresponds to a quarter), the true latent asset return correlation is \( \rho_b = 0.05 \), the quarterly probability of default is set equal to 0.2%, and the plots are based on 50,000 simulations for each AR(1) and i.i.d. time series.
obligors so that the idiosyncratic risk is diversified away. We further assume that loss given default is 50% so that we can compute the VaR over the 50,000 simulated paths of 4-quarter default time series conditional on the given history of the last 80 quarters of default time series. We then compute the unconditional VaR on different levels for each of the previously discussed estimators. The results are displayed in Table 1 for AR(1) underlying dynamics.

<table>
<thead>
<tr>
<th>$\varrho_b$</th>
<th>true value</th>
<th>MoM unadjusted</th>
<th>MoM adjusted ($k = 5$)</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EL</td>
<td>3.99</td>
<td>3.99 (+0.05%)</td>
<td>3.99 (+0.09%)</td>
<td>4.00 (+0.10%)</td>
</tr>
<tr>
<td>$\text{VaR}_{0.95}$</td>
<td>6.98</td>
<td>6.82 (-2.33%)</td>
<td>6.99 (+0.14%)</td>
<td>6.87 (-1.70%)</td>
</tr>
<tr>
<td>$\text{VaR}_{0.99}$</td>
<td>9.26</td>
<td>8.94 (-3.44%)</td>
<td>9.31 (+0.54%)</td>
<td>9.03 (-2.47%)</td>
</tr>
<tr>
<td>$\text{VaR}_{0.999}$</td>
<td>12.71</td>
<td>12.15 (-4.45%)</td>
<td>12.87 (+1.21%)</td>
<td>12.31 (-3.13%)</td>
</tr>
<tr>
<td>$\text{VaR}_{0.9999}$</td>
<td>13.80</td>
<td>13.15 (-4.68%)</td>
<td>13.99 (+1.41%)</td>
<td>13.34 (-3.26%)</td>
</tr>
</tbody>
</table>

Table 1: Expected loss (EL) and Value-at-Risk (VaR) in an example of a diversified $1 bn loan portfolio. The true values are computed using the true latent asset return correlation $\varrho_b = 0.05$ (upper half) or $\varrho_b = 0.10$ (lower half) and parameter 0.7 of the AR(1) process. All numbers are in million $, with percentage deviations compared to the true value. As in the example of Figure 4, the quarterly probability of default is set equal to 0.2%.

In line with our expectations, we find that estimated VaR$\alpha$ numbers are closer to their true values for lower levels $\alpha$ and lower asset return correlation $\varrho_b$. Generally, our adjusted estimator (adjustment of order $k = 5$) leads to estimated VaR closest to the true values, better than the MLE and better than the unadjusted moment estimator. Interestingly, we find that VaR estimations are slightly too high when using our adjusted estimator in the case of $\varrho_b = 0.05$. This is because both the latent asset return correlation and the coefficient of the AR(1) process are estimated. If one chooses the correct coefficient for the AR(1) process instead of estimating it, we obtain for $\text{VaR}_{0.999}$ when $\varrho_b = 0.05$, the values $11.95m (-5.96\%)$ compared to the true value) for the unadjusted MoM, $12.65m (-0.48\%)$ for the adjusted MoM, and $12.12m (-4.68\%)$ for the MLE. In this example, we observe that, for VaR risk metrics, the estimation of the autocorrelation mitigates parts of the bias issues of the latent asset return estimators, but still VaR numbers estimated by MLE and MoM are too low and improved when our proposed adjustment is applied to MoM. The low default probability (0.2% per quarter) means that VaR$\alpha$ numbers and the impact on misestimated correlation are relatively small, but even under these choices, the adjusted MoM performs clearly better than the other estimators. The adjustment is particularly beneficial for higher levels of asset return correlation: already for $\varrho_b = 0.10$, $\text{VaR}_{0.999}$ is underestimated by around 8% and 4%

We also analyzed underlying dynamics driven by i.i.d. observations, rather than an AR(1) process, and found, as expected, that the deviations in VaR from the different estimators are typically small (less than 2%) because the bias from autocorrelation is not present in this case.
by unadjusted MoM and MLE, respectively, while our adjusted MoM estimator provides essentially unbiased VaR numbers.

5 Conclusion

Starting with the observation that default rate time series often exhibit autocorrelation, we examined how to relax the assumption of serial independence in classical estimators used in credit risk modelling. This led us to study convergence and distributional properties of the sample mean of general autocorrelated time series. Under mild assumptions, the sample mean still converges, but more slowly than in the case of serial independence. For finite time series, the slower convergence causes a bigger bias of classical estimators when autocorrelation is present. Thus, we constructed an estimator that includes correction terms taking the autocorrelation and shortness of the observed time series into account. Applied to credit risk modelling, we found that our estimator removes a big part of the downward bias that classical estimators for the latent asset return correlation have. The explicit formula of our estimator, which does not depend on distributional assumptions, makes the estimator easily tractable and readily available for industrial applications. Its implementation helps determine more accurate values of aggregate risk metrics such as Value-at-Risk, which crucially depend on good estimations for the latent asset return correlation.
A Formula for new estimator of return correlation

For practical purposes, we summarize our estimator for latent intra-segment asset return correlation in the case of a first-order correction term, which typically captures a big part of the bias of the classical estimator; compare Figure 1. Formulas for higher-order correction terms and inter-segment correlation can be found in Section 4.1.

**Input:**
default rate times series \((p_t)_{t=1,...,T}\) for a bucket in a Merton model (see Section 2)

**Estimator:**
\[ \hat{\varrho}_2 = \hat{\varrho}_1 + \frac{g''(\hat{\varrho}_1)}{T(g'(\hat{\varrho}_1))^3} \left( \frac{\alpha_0}{2} + (1 - 1/T)\alpha_1 \right) \]

**Components in estimator:**
- classical estimator:
  \[ \hat{\varrho}_1 = g^{-1}\left( \frac{1}{T} \sum_{t=1}^{T} p_t^2 \right) \]

  where
  \[ g(\varrho) = \Phi_2\left( \Phi_{(-1)}\left( \frac{1}{T} \sum_{t=1}^{T} p_t \right), \Phi_{(-1)}\left( \frac{1}{T} \sum_{t=1}^{T} \bar{p}_t \right); \varrho \right) \]

  with \( \Phi_2(.,.; \varrho) \) denoting the bivariate normal cumulative distribution function with correlation \( \varrho \).

- derivatives appearing in correction term:
  \[ g'(\hat{\varrho}_1) = \frac{1}{2\pi \sqrt{1 - \hat{\varrho}_1^2}} \exp\left( - \frac{s^2}{1 + \hat{\varrho}_1} \right), \]
  \[ g''(\hat{\varrho}_1) = \frac{s^2 + \hat{\varrho}_1(1 - 2s^2) + s^2 \hat{\varrho}_1^2 - \hat{\varrho}_1^3}{2\pi (1 - \hat{\varrho}_1^2)^{5/2}} \exp\left( - \frac{s^2}{1 + \hat{\varrho}_1} \right) \]

  with \( s = \Phi_{(-1)}\left( \frac{1}{T} \sum_{t=1}^{T} p_t \right) \).

- sample variance and lag-1 sample autocovariance:
  \[ \alpha_0 = \frac{1}{T} \sum_{t=1}^{T} (p_t^2 - \bar{\mu})^2, \quad \alpha_1 = \frac{1}{T} \sum_{t=2}^{T} (p_t^2 - \bar{\mu})(p_{t-1}^2 - \bar{\mu}), \]

  where \( \bar{\mu} = \frac{1}{T} \sum_{t=1}^{T} p_t^2 \).
B Proof of Theorem 3.3

By Taylor’s theorem, we can write
\[
\tilde{g}\left(\frac{1}{T} \sum_{t=1}^{T} Z_t\right) = \tilde{g}(\mu) + \tilde{g}'(\mu)\left(\frac{1}{T} \sum_{t=1}^{T} Z_t - \mu\right) + \frac{\tilde{g}''(\mu)}{2} \left(\frac{1}{T} \sum_{t=1}^{T} Z_t - \mu\right)^2 \\
+ \frac{\tilde{g}'''(\xi)}{6} \left(\frac{1}{T} \sum_{t=1}^{T} Z_t - \mu\right)^3
\]
for some \(\xi\) between \(\mu\) and \(\frac{1}{T} \sum_{t=1}^{T} Z_t\). Taking expectations and rearranging terms yield
\[
\tilde{g}(\mu) - E\left[\tilde{g}\left(\frac{1}{T} \sum_{t=1}^{T} Z_t\right)\right] + \frac{\tilde{g}''(\mu)}{2} \text{Var}\left(\frac{1}{T} \sum_{t=1}^{T} Z_t\right) = -E\left[\frac{\tilde{g}'''(\xi)}{6} \left(\frac{1}{T} \sum_{t=1}^{T} Z_t - \mu\right)^3\right], \tag{18}
\]
using that \(E[Z_t] = \mu\) for all \(t\) by stationarity. We now analyze the different terms in (18). First, we note \(\tilde{g}(\mu) = \theta\) and compute
\[
\tilde{g}''(\mu) = \left(\frac{1}{g'(g^{-1}(\mu))}\right)' = -\frac{g''(g^{-1}(\mu))(g^{-1})'(\mu)}{(g'(g^{-1}(\mu)))^2} = -\frac{g''(g^{-1}(\mu))}{(g'(g^{-1}(\mu)))^3} = -\frac{g''(\theta)}{(g'(\theta))^3}. \tag{19}
\]
Similarly to (9), we can write the variance term as
\[
\text{Var}\left(\frac{1}{T} \sum_{t=1}^{T} Z_t\right) = \frac{1}{T} \text{Var}(Z_1) + \frac{2}{T^2} \sum_{\ell=1}^{T-1} (T - \ell) \text{Cov}(Z_1, Z_{1+\ell}). \tag{20}
\]
Finally, we apply Hölder’s inequality with \(p = 4/3\) and \(q = 4\) (which satisfy \(1/p + 1/q = 1\)) to obtain
\[
E\left[\left|\tilde{g}'''(\xi)\left(\frac{1}{T} \sum_{t=1}^{T} Z_t - \mu\right)^3\right|^p\right] \leq E\left[|\tilde{g}'''(\xi)|^q\right]^{1/q} E\left[\left(\frac{1}{T} \sum_{t=1}^{T} Z_t - \mu\right)^3\right]^{p'}^{1/p} \\
\leq E\left[\left(\frac{1}{T} \sum_{t=1}^{T} Z_t - \mu\right)^4\right]^{3/4} E\left[\left(\frac{1}{T} \sum_{t=1}^{T} Z_t - \mu\right)^3\right]^{3/4}
\]
which concludes the proof in light of (18)–(20). \(\Box\)

Declaration of interest

The research was done while Marcus Wunsch was employed by UBS and Christoph Frei was working on projects at UBS while on sabbatical from the University of Alberta. The views expressed in the paper are those of the authors and do not necessarily reflect views of UBS AG, its subsidiaries or affiliates.
Acknowledgements

We would like to thank Paul Embrechts, Michael Gordy, Gregor Heyne, Hans Rudolf Künsch, Jonathan Wendin and two anonymous referees as well as participants of a seminar at ETH Zurich for helpful comments. Christoph Frei gratefully acknowledges support for research projects by the Natural Sciences and Engineering Council of Canada under grant RGPIN/402585-2011.

References


