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

Christoph Frei (cfrei@ualberta.ca)[1] and Marcus Wunsch (marcus.wunsch@ubs.com)[2]

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

This version: November 7, 2017
(Journal of Credit Risk 14 (2018), 1–29)

Abstract

In credit risk modelling, method-of-moment approaches are popular to estimate latent asset return correlations within and between rating buckets. However, the autocorrelation that is often present in time series of default rates leads to systematically too low estimations. We propose a new estimator that adjusts for the problems of this autocorrelation and the shortness of the time series, thus eliminating a significant portion of the bias observed in classical estimators. The adjustment is based on convergence and approximation results for general autocorrelated time series, and is easily implementable and nonparametric.

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

1 Introduction

In both practice and academia, the credit risk of a loan portfolio is frequently modelled using a structural approach that generalizes Merton (1974) to multiple firms. Obligors in the portfolio are grouped into homogeneous buckets according to their client segments and ratings (thus the term, "rating buckets"). After Merton, the assumption is that the obligor's latent asset return process will determine their default: if the return over the model horizon is below some threshold, the obligor defaults. This return process is driven by systematic and idiosyncratic risk factors that thus introduce a dependence structure among the obligors' returns and, in turn, 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 typically estimated via the method-of-moment approach or the maximum likelihood method, based on historical default rates within each bucket.

In the classical literature (see Bluhm et al. (2010) or McNeil et al. (2015) 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. Therefore, default rates for a given rating bucket will be serially independent. In practice, however, the data exhibit autocorrelation in the default rates [1] and thus in the latent asset returns. Two reasons account 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; default rates thus reflect the behaviour of economic factors over time, which exhibit autocorrelation. Through-the-cycle parameters are typically chosen in these instances because they improve the stability of credit risk measures across economic cycles. For instance, Basel's advanced internal rating-based approach includes the statement "Although the time horizon used in PD [probability of default] estimation is one year ... banks must use a longer time horizon in assigning ratings." [2] A current proposal of the Basel Committee explicitly suggests that "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." [3] Moreover, in 2011, Moody's Analytics introduced *through-the-cycle* expected default frequency (EDF) to complement its *point-in-time* EDF, and thereby reduce short-term volatility from the credit cycle.[4] The risk management industry and academic literature have widely discussed the meaning and (dis)advantages of ratings as point-in-time versus through-the-cycle credit indicators; see Cantor and Mann (2006); Carey and Hrycay (2001); Gordy and Howells (2006); Heitfield (2004); Löffler (2004).

Even when attempting to use point-in-time ratings that reflect the current economic situation, autocorrelation can still be present because changes in the credit quality of obligors may not be immediately reflected in their ratings. This delay affects default rates for the rating buckets, making them dependent on the economic situation, and subsequently autocorrelated. Ratings that slowly adjust are a well documented and explained phenomenon; see Altman and Rijken (2006), and Löffler (2005). Regardless of what accounts for the observed autocorrelation in default rates, when these time series are inputs in credit risk estimators, their autocorrelation must be considered.

The goal of this paper is twofold. First, we analyze how autocorrelation affects method-of-moment estimators commonly used in industry to determine the latent asset return correlation. While we find that the stationarity and summability of auto-co-moments are sufficient for the estimator to converge to the true asset return correlation when the length of the time series goes to infinity, we show that the convergence rate is much slower than in the case of independent and identically distributed (i.i.d.) observations. Second, we propose a new estimator that includes correction terms accounting for the autocorrelation and shortness of

[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.

[2] Paragraph 414 in Basel Committee on Banking Supervision. International convergence of capital measures and capital standards, 2006.

[3] 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/webinars-on-demand/2011/through-cycle-edf

the observed time series. Our estimator, for which we also provide confidence intervals, has a simple explicit form. Interestingly, the correction terms are positive when positive autocorrelation is present. This shows that asset return correlation and therefore the values of aggregate risk metrics are typically underestimated by classical estimators, which are based on the method of moments and assume serial independence. Our proposed adjustment vanishes as the length of the time series goes to infinity; it becomes more dominant as the time series shortens. 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.

Gordy and Heitfield (2010) have described the estimation bias in the asset return correlation in cases of serially independent asset returns. In this paper, we allow for autocorrelation and provide a remedy for the intrinsic bias that classical estimators have by proposing a new estimator. 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 the law of large numbers and a 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, from which we derive the correction terms included in our new estimator.

Maximum likelihood estimators are an alternative to the method of moments 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 (2007); see also McNeil and Wendin (2006); Wendin (2006) for related results and additional details. They proposed applying computational Bayesian estimation (Gibbs sampling) to the maximum likelihood estimation of a class of generalized linear mixed models (GLMMs) being used to model default rates. In their approach, computational Bayesian estimation allows for a variety of model specifications and priors, so that different information sources and model structures can be used. In contrast, our approach has the following advantages:

1. Most importantly, it leads to an easily implementable, explicit formula that directly shows the impact of autocorrelation on the estimation value. Thus, it is very suitable for industrial applications.
2. Our method is stable, since it does not require any distributional properties of the underlying time series. Gibbs sampling, on the other hand, does demand assumptions about prior distribution and the dynamics of the time series (for example, an AR(1) process).
3. 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.

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. These results are presented for a well diversified portfolio while we explain in Section 4.2 how our approach can be adjusted to account for obligor specific risk. In Sections 4.3 and 4.4, we compare our estimator with maximum likelihood estimators, also considering the implications of both types of estimators on a loss risk metrics. In Section 5, we conclude. In the appendix to this paper, we summarize how the formula for our estimator can be applied in practice, and provide the proof of our main result.

2 Revisiting a structural default model

In this section, we recall method-of-moment estimators for structural credit risk models of the Merton type. Method-of-moment estimators were introduced in this context by Gordy (2000) and Nagpal and Bahar (2001) and later refined by Frey and McNeil (2003), and Bluhm and Overbeck (2003), among others. We refer to Bluhm et al. (2010) 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 approach generalizing Merton (1974) to multiple firms. We group obligors in the portfolio into homogeneous buckets. We assume here that the portfolio is infinitely fine grained in the sense that the idiosyncratic risk of obligors is negligible on the bucket level. In Section 4.2, we will explain how the approach can be adjusted when idiosyncratic risk is present. 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 ϵ_i is a standard normally distributed random variable (the obligor i's idiosyncratic component) independent of the other ϵ_i and Y_b. For two obligors i and $\tilde{i}$ in different buckets b and $\tilde{b}$, we write

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

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

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

so that $c_b = \Phi^{-1}(p_b)$. Using the independence of ϵ_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 ϱ_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 large number of obligors so that the idiosyncratic risk is diversified away at the bucket level. We assume that we are given the default rate time series $(p_b(Y_{b,t}))_{t=1,\dots,T}$ and $(p_{\tilde{b}}(Y_{\tilde{b},t}))_{t=1,\dots,T}$ for the two buckets, which correspond to time series $(Y_{b,t})_{t=1,\dots,T}$ and $(Y_{\tilde{b},t})_{t=1,\dots,T}$ of systematic factors, reflecting economic conditions. We further assume that $(p_b(Y_{b,t})^2)_{t=1,\dots,T}$ and $(p_b(Y_{b,t})p_{\tilde{b}}(Y_{\tilde{b},t}))_{t=1,\dots,T}$ are stationary, but importantly, they do not need to be serially independent so that they can exhibit autocorrelation, as we observe in practice.

Proposition 2.5.9 in Bluhm et al. (2010) shows that

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

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

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}(p_b(Y_{b,t}))^2 = \Phi_2\big(\Phi^{(-1)}(p_b), \Phi^{(-1)}(p_b); \varrho_b\big) \quad \text{a.s.}$$

by Proposition 3.1 below so that we obtain the approximation

$$\frac{1}{T}\sum_{t=1}^{T}(p_b(Y_{b,t}))^2 \approx \Phi_2\left(\Phi^{(-1)}\left(\frac{1}{T}\sum_{t=1}^{T}p_b(Y_{b,t})\right), \Phi^{(-1)}\left(\frac{1}{T}\sum_{t=1}^{T}p_b(Y_{b,t})\right); \varrho_b\right), \tag{2}$$

using here the assumption that the idiosyncratic risk is diversified away on the bucket level. We consider the function

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

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 ω. Restricting g_b to nonnegative ϱ_b and fixing a scenario ω with $y_{b,t} = Y_{b,t}(\omega)$, it can be shown that g_b is invertible[5] so that we obtain an estimator

$$\hat{\varrho}_{b,1} = g_b^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}(p_b(y_{b,t}))^2\right). \tag{3}$$

From Gordy and Heitfield (2010), it is known that $\hat{\varrho}_{b,1}$ is a biased estimator of ϱ_b because of the finite length of the time series $(p_b(Y_{b,t}))_{t=1,\dots,T}$, 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 (2003), we can approximate

$$\frac{1}{T}\sum_{t=1}^{T}p_b(Y_{b,t})p_{\tilde{b}}(Y_{\tilde{b},t}) \approx \Phi_2\left(\Phi^{(-1)}\left(\frac{1}{T}\sum_{t=1}^{T}p_b(Y_{b,t})\right), \Phi^{(-1)}\left(\frac{1}{T}\sum_{t=1}^{T}p_{\tilde{b}}(Y_{\tilde{b},t})\right); \varrho_{b,\tilde{b}}\right),$$

[5] g_b is strictly increasing as it will follow later from (11) and $g_b(0) = \big(\frac{1}{T}\sum_{t=1}^{T}p_b(y_{b,t})\big)^2 \le \frac{1}{T}\sum_{t=1}^{T}p_b(y_{b,t})^2$ by Jensen's inequality and $g_b(1) = \frac{1}{T}\sum_{t=1}^{T}p_b(y_{b,t}) \ge \frac{1}{T}\sum_{t=1}^{T}p_b(y_{b,t})^2$ since $p_b(y_{b,t}) \in [0,1]$.

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

$$\hat{\varrho}_{b,\tilde{b},1} = g_{b,\tilde{b}}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} p_b(y_{b,t})p_{\tilde{b}}(y_{\tilde{b},t})\right), \tag{4}$$

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

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

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 ω.

3 Convergence results for autocorrelated time series

In this section, we analyze the behaviour of the sample mean $\frac{1}{T}\sum_{t=1}^{T} Z_t$ of an autocorrelated time series $(Z_t)_{t=-\infty}^{t=\infty}$. 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

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

The next result follows from an application of the ergodic theorem; see, for example, Theorem 19.1 in Greene (2008).

Proposition 3.1. *Under assumptions A1 and A2, the sample mean of the time series* $(Z_t)_{t=-\infty}^{t=\infty}$ *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-co-kurtosis 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 $|\rho| < 1$,[6] $\mu \in \mathbb{R}$, $\sigma > 0$, and independent, standard normally distributed ϵ_t. This process can be demeaned by subtracting the mean μ, i.e., the process $Z_t = X_t - \mu$ has zero mean. 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 co-moments of Z_s and Z_t are finite. In particular, if we define an auto-co-kurtosis $\kappa_{s,t} = \kappa_{s,t}^{(2,2)}$ as[7]

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

then

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

for integers s and t. Similar (summable) expressions hold for the auto-co-kurtoses $\kappa_{s,t}^{(1,3)}$ 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. We can generalize this application to the situation where the innovations ϵ_t of the AR(1) process are not normally distributed, provided that their kurtosis $E[\epsilon_t^4]$ is finite. Indeed, the auto-co-kurtosis $\kappa_{s,t}$ of the AR(1) process is finite if and only if $E[\epsilon_t^4]$ is finite. Thus, if $E[\epsilon_t^4]$ is finite when ϵ_t is not necessarily normally distributed, we can still apply Proposition 3.1 to obtain almost sure convergence of the second sample moment of the AR(1) process.

Generalizing from AR(1) processes, 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 Hamilton (1994).

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

[6] In this case, the AR(1) process is called wide-sense stationary.

[7] Note that in addition to this version of the auto-co-kurtosis with powers $(2,2)$, there are two other versions with powers $(1,3)$ and $(3,1)$.

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

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

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

$$\sum_{k=0}^{\infty} E[r_{t,k}^2] = \sum_{k=0}^{\infty} \rho^{2k} E[\epsilon_{t-k}^2] = \frac{1}{1-\rho^2} < \infty.$$

Remark 3.2. *By the convergence theorem for backward martingales (see, for example, Section 5.6 in Durrett (2010)), we always have*

$$E[Z_t|Z_{t-k}, Z_{t-k-1}, \ldots] \overset{k\to\infty}{\to} E[Z_t|\mathcal{F}_{-\infty}] \quad \text{in } L^1 \text{ and almost surely,}$$

with σ*-algebra* $\mathcal{F}_{-\infty} := \bigcap_{k=0}^{\infty} \sigma(Z_{t-k}, Z_{t-k-1}, \ldots)$. *In particular, this implies* $r_{t,k} \to 0$ *in* L^1 *and almost surely as* $k \to \infty$. *Assumptions A3 and A4 reinforce these general convergence properties by imposing that the sequence* $E[Z_t|Z_{t-k}, Z_{t-k-1}, \ldots]$ *also converges in* L^2 *as* $k \to \infty$ *and that* $\sum_{k=0}^{\infty} E[r_{t,k}^2]$ *is finite for fixed* t.

Gordin's central limit theorem (see Theorem 19.4 in Greene (2008)) says that under assumptions A1–A4

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t \xrightarrow{T\to\infty} \mathcal{N}\Big(0, \sum_{t=-\infty}^{\infty} \gamma_{0,t}\Big) \text{ in distribution,} \tag{5}$$

where $\mathcal{N}\big(0, \sum_{t=-\infty}^{\infty} \gamma_{0,t}\big)$ is the normal distribution with mean zero and variance $\sum_{t=-\infty}^{\infty} \gamma_{0,t}$.

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 θ is a parameter to be estimated, μ is the unknown mean of the stationary time series $(Z_t)_{t=1,\dots,T}$. We assume that g is three times continuously differentiable and invertible with inverse $\tilde{g} = g^{-1}$.

A natural estimator for θ is the moment estimator

$$\hat{\theta}_1 = \tilde{g}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\right). \tag{6}$$

By continuity of $\tilde{g}$ and the law of large numbers (see Proposition 3.1), $\hat{\theta}_1$ converges almost surely to θ 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 the appendix to this paper, allows us to improve the estimation.

Theorem 3.3. *Under assumption A1, there exists a random variable ξ with values between μ and $\frac{1}{T}\sum_{t=1}^{T} Z_t$ such that*

$$\left|\theta - E\left[\tilde{g}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\right)\right] - \frac{g''(\theta)}{2T(g'(\theta))^3}\mathrm{Var}(Z_1) - \frac{g''(\theta)}{T^2(g'(\theta))^3}\sum_{\ell=1}^{T-1}(T-\ell)\mathrm{Cov}(Z_1, Z_{1+\ell})\right|$$
$$\leq \frac{1}{6}E\big[(\tilde{g}'''(\xi))^4\big]^{1/4} E\left[\left(\frac{1}{T}\sum_{t=1}^{T} Z_t - \mu\right)^4\right]^{3/4}.$$

The theorem says that θ can be approximated by

$$E\left[\tilde{g}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\right)\right] + \frac{g''(\theta)}{2T(g'(\theta))^3}\mathrm{Var}(Z_1) + \frac{g''(\theta)}{T^2(g'(\theta))^3}\sum_{\ell=1}^{T-1}(T-\ell)\mathrm{Cov}(Z_1, Z_{1+\ell}), \tag{7}$$

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: θ is replaced by the estimator $\hat{\theta}_1 = \tilde{g}\big(\frac{1}{T}\sum_{t=1}^{T} Z_t\big)$, and $\mathrm{Var}(Z_1)$ and $\mathrm{Cov}(Z_1, Z_{1+\ell})$ are replaced by sample variance and sample autocovariances. This gives us a new estimator

$$\hat{\theta}_2 = \tilde{g}(\bar{\mu}) + \frac{g''(\tilde{g}(\bar{\mu}))}{2T(g'(\tilde{g}(\bar{\mu})))^3}\alpha_0 + \frac{g''(\tilde{g}(\bar{\mu}))}{T(g'(\tilde{g}(\bar{\mu})))^3}\sum_{\ell=1}^{k}(1-\ell/T)\alpha_\ell, \tag{8}$$

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

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

is the lag-ℓ 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 (2017).

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

$$
\begin{aligned}
\mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\right) &= \frac{1}{T}\mathrm{Var}\left(\sum_{t=1}^{T} Z_t\right) = \frac{1}{T}\sum_{i,j=1}^{T}\mathrm{Cov}(Z_i, Z_j) \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathrm{Var}(Z_t) + \frac{2}{T}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\mathrm{Cov}(Z_t, Z_{t+\ell}) \\
&= \mathrm{Var}(Z_1) + \frac{2}{T}\sum_{\ell=1}^{T-1}(T-\ell)\mathrm{Cov}(Z_1, Z_{1+\ell}) \qquad (9) \\
&\approx \alpha_0 + \frac{2}{T}\sum_{\ell=1}^{k}(T-\ell)\alpha_\ell.
\end{aligned}
$$

Applying Gordin's central limit theorem (5) and using the delta method, we can approximate a $p\,\%$ confidence interval for $\hat{\theta}_2$ by

$$
\left[\hat{\theta}_2 - \frac{q|\tilde{g}'(\bar{\mu})|}{\sqrt{T}}\sqrt{\alpha_0 + 2\sum_{\ell=1}^{k}(1-\ell/T)\alpha_\ell}, \hat{\theta}_2 + \frac{q|\tilde{g}'(\bar{\mu})|}{\sqrt{T}}\sqrt{\alpha_0 + 2\sum_{\ell=1}^{k}(1-\ell/T)\alpha_\ell}\right] \qquad (10)
$$

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 (1996), 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_ℓ *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* α_k *to the weight 0 for* α_{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 (1996), 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-moment estimators from Section 2 can be improved by using the results of Section 3.2. In Section 4.2, we explain how an additional adjustment can be made to the estimator in the presence of idiosyncratic risk on the bucket level. In Section 4.3, we give a performance comparison of the original and adjusted method-of-moment estimators with a maximum likelihood estimator (MLE). Section 4.4 gives an example of 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 large 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

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

as realizations of (6) with $\tilde{g} = g_b^{(-1)}$, $\hat{\varrho}_1 = \hat{\varrho}_{b,1}$, $Z_t = (p_b(Y_{b,t}))^2$ and $\tilde{g} = g_{b,\tilde{b}}^{(-1)}$, $\hat{\varrho}_1 = \hat{\varrho}_{b,\tilde{b},1}$, $Z_t = p_b(Y_{b,t})p_{\tilde{b}}(Y_{\tilde{b},t})$, respectively. Since, in either case, the inverse of $\tilde{g}$ is $g(\varrho_1) = \Phi_2(s,t;\varrho_1)$, we calculate the derivatives

$$\frac{\partial}{\partial\varrho_1}\Phi_2(s,t;\varrho_1) = \phi_2(s,t;\varrho_1) = \frac{1}{2\pi\sqrt{1-\varrho_1^2}}\exp\left(-\frac{s^2/2-\varrho_1 st+t^2/2}{1-\varrho_1^2}\right), \tag{11}$$

$$\frac{\partial^2}{(\partial\varrho_1)^2}\Phi_2(s,t;\varrho_1) = \frac{st+\varrho_1(1-s^2-t^2)+st\varrho_1^2-\varrho_1^3}{2\pi(1-\varrho_1^2)^{5/2}}\exp\left(-\frac{s^2/2-\varrho_1 st+t^2/2}{1-\varrho_1^2}\right). \tag{12}$$

From Section 3.2, we can find estimated correlations $\hat{\varrho}_{b,2}$ and $\hat{\varrho}_{b,\tilde{b},2}$, taking the short length and autocorrelation of the time series into account, by

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

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

where

- $\hat{\varrho}_{b,1}$ and $\hat{\varrho}_{b,\tilde{b},1}$ are the estimations from (3) and (4);
- $g'(\hat{\varrho}_{b,1})$ and $g''(\hat{\varrho}_{b,1})$ equal to (11) and (12) with $\varrho_1 = \hat{\varrho}_{b,1}$ and evaluated at $s = t = \Phi^{(-1)}\big(\frac{1}{T}\sum_{t=1}^{T} p_b(y_{b,t})\big)$; analogously, $g'(\hat{\varrho}_{b,\tilde{b},1})$ and $g''(\hat{\varrho}_{b,\tilde{b},1})$ equal to (11) and (12) with $\varrho_1 = \hat{\varrho}_{b,\tilde{b},1}$ and evaluated at $s = \Phi^{(-1)}\big(\frac{1}{T}\sum_{t=1}^{T} p_b(y_{b,t})\big)$ and $t = \Phi^{(-1)}\big(\frac{1}{T}\sum_{t=1}^{T} p_{\tilde{b}}(y_{\tilde{b},t})\big)$;
- $\alpha_{b,\ell}$ and $\alpha_{b,\tilde{b},\ell}$ are the lag-ℓ sample autocovariances of the time series $\big((p_b(y_{b,t}))^2\big)_{t=1,\dots,T}$ and $(p_b(y_{b,t})p_{\tilde{b}}(y_{\tilde{b},t}))_{t=1,\dots,T}$, respectively, i.e.,

$$\alpha_{b,\ell} = \frac{1}{T}\sum_{t=1+\ell}^{T}\big((p_b(y_{b,t}))^2-\bar{\mu}_b\big)\big((p_b(y_{b,t-\ell}))^2-\bar{\mu}_b\big), \tag{13}$$

$$\alpha_{b,\tilde{b},\ell} = \frac{1}{T}\sum_{t=1+\ell}^{T}\big(p_b(y_{b,t})p_{\tilde{b}}(y_{\tilde{b},t})-\bar{\mu}_{b,\tilde{b}}\big)\big(p_b(y_{b,t-\ell})p_{\tilde{b}}(y_{\tilde{b},t-\ell})-\bar{\mu}_{b,\tilde{b}}\big),$$

  where $\bar{\mu}_b = \frac{1}{T}\sum_{t=1}^{T}(p_b(y_{b,t}))^2$ and $\bar{\mu}_{b,\tilde{b}} = \frac{1}{T}\sum_{t=1}^{T} p_b(y_{b,t})p_{\tilde{b}}(y_{\tilde{b},t})$.

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 $\hat{\varrho}_{b,1}$ is much larger 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 $\varrho_b = 0.05$, 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 $\hat{\varrho}_{b,1}$, 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 $\hat{\varrho}_{b,1}$ because of its appearance as argument in Φ_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 refrain from using such an estimator because we observe that the additionally gained precision is small for the estimation, the notation is cumbersome and there is no explicit expression for derivatives analogous to (11) and (12). However, we used such an idea to get good approximated confidence intervals in Figure 2. Indeed, we use confidence intervals $[\hat{\varrho}_{b,2} - s_b, \hat{\varrho}_{b,2} + s_b]$ where we choose s_b differently than $\frac{q|\tilde{g}'(\bar{\mu})|}{\sqrt{T}}\sqrt{\alpha_0 + 2\sum_{\ell=1}^{k}(1-\ell/T)\alpha_\ell}$ in (10). From

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

and Footnote 5, there exists a function h with $h\big(\frac{1}{T}\sum_{t=1}^{T}(p_b(y_{b,t}))^2, \frac{1}{T}\sum_{t=1}^{T}p_b(y_{b,t})\big) = \hat{\varrho}_b$. 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

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

where

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

$$\gamma_{b,\ell} = \frac{1}{T-1-\ell}\sum_{t=1+\ell}^{T}\big(p_b(y_{b,t}) - \bar{\nu}_b\big)\big(p_b(y_{b,t-\ell})^2 - \bar{\mu}_b\big)$$

with $\bar{\nu}_b = \frac{1}{T}\sum_{t=1}^{T}p_b(y_{b,t})$ 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.

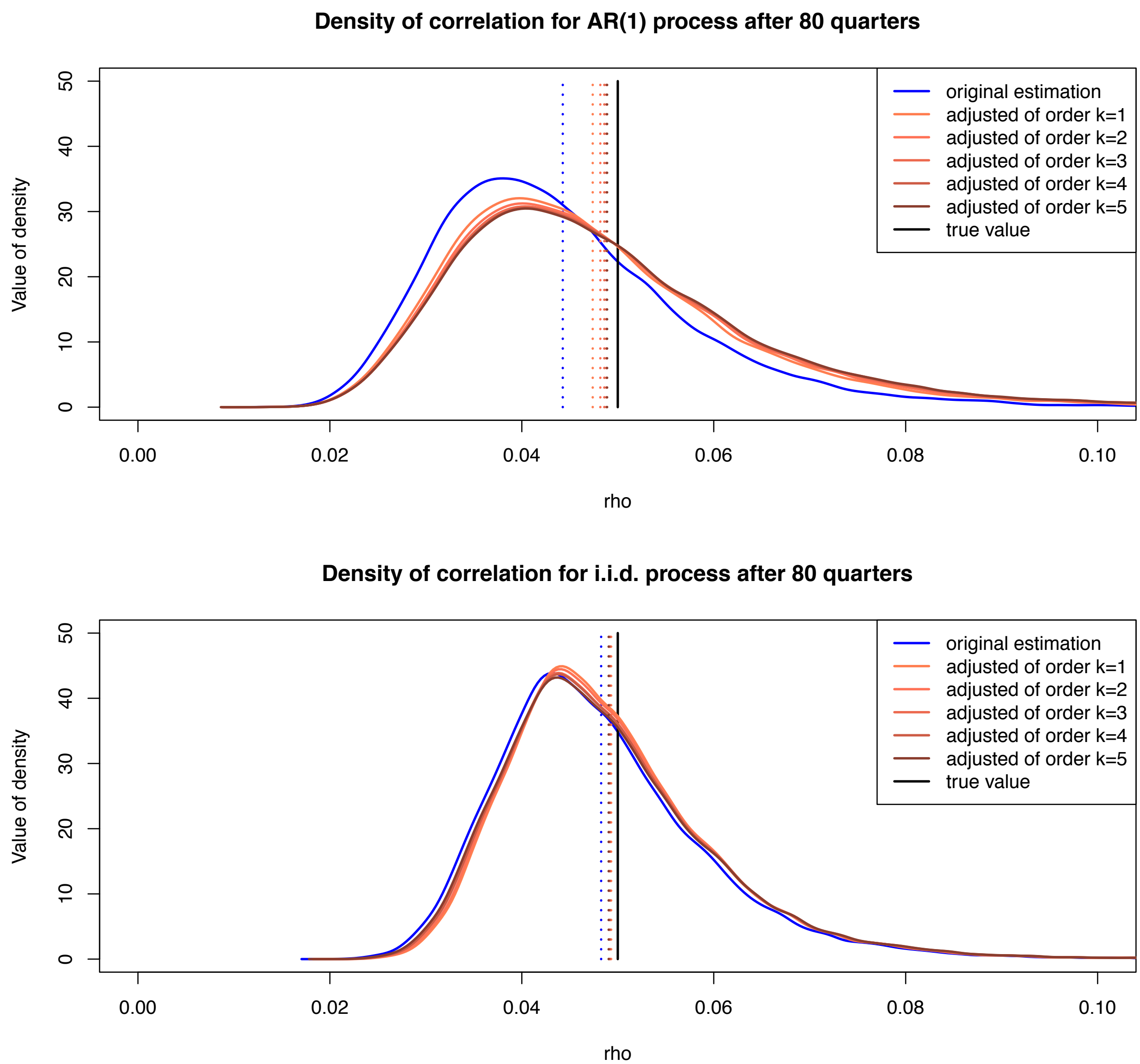

Figure 1: Comparison of distributions of the original correlation estimation and the adjusted correlation estimation of different orders $k = 1, \ldots, 5$. The underlying factors are simulated based on an AR(1) process with coefficient 0.7 (upper panel) and i.i.d. observations (lower panel). The solid 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 $\varrho_b = 0.05$, 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.

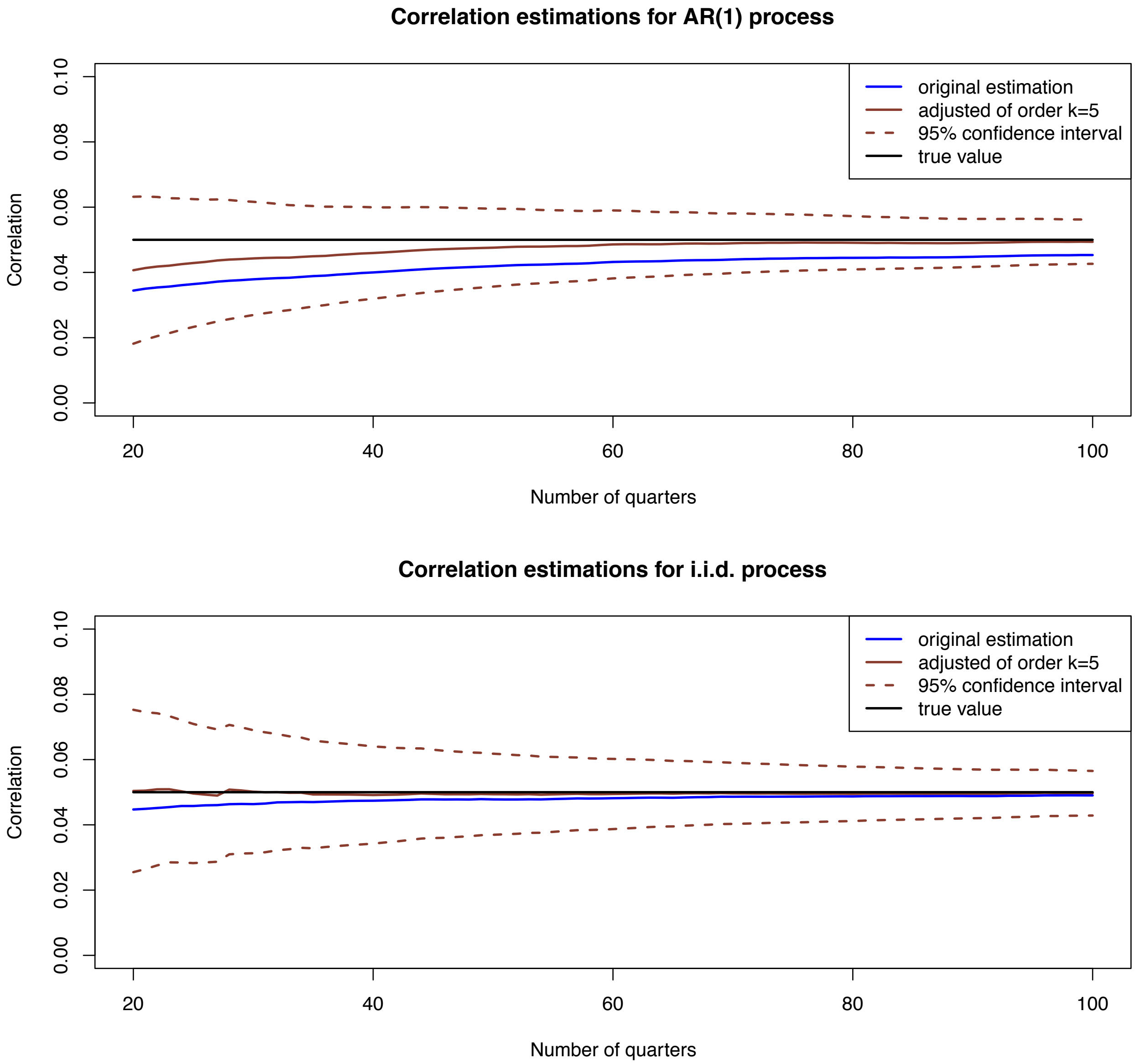

Figure 2: Comparison of the original correlation estimation and the adjusted correlation estimation of orders $k = 5$ for different lengths of the time series. The underlying factors are simulated based on an AR(1) process with coefficient 0.7 (upper panel) and i.i.d. observations (lower panel). The 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 $\varrho_b = 0.05$, 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.

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* ϱ_1 *the sum of squared errors*

$$\sum_{b\in B} w_b \left[\varrho_1 - g_b^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} (p_b(y_{b,t}))^2 \right) \right]^2 .$$

The optimal ϱ_1 *is given by the first-order condition* $\sum_{b\in B} w_b \big[\varrho_1 - g_b^{-1}\big(\frac{1}{T}\sum_{t=1}^{T}(p_b(y_{b,t}))^2\big)\big] = 0$*, hence* $\varrho_1 = \sum_{b\in B} w_b\, g_b^{-1}\big(\frac{1}{T}\sum_{t=1}^{T}(p_b(y_{b,t}))^2\big)$*. 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.*

4.2 Accounting for idiosyncratic risk

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. In this section, we 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 $\hat{\varrho}_b$ is defined by

$$\Phi_2\left(\Phi^{(-1)}\left(\frac{1}{T}\sum_{t=1}^{T} x_t \right), \Phi^{(-1)}\left(\frac{1}{T}\sum_{t=1}^{T} x_t \right); \hat{\varrho}_b \right) = \frac{1}{T}\sum_{t=1}^{T} x_t^2. \tag{14}$$

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 larger 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 Φ_2 is increasing in $\hat{\varrho}_b$ (compare (11)), this will typically lead to a larger estimation $\hat{\varrho}_b$.

This additional bias resulting from the finite number N of obligors can be partially corrected. Indeed, writing $X_t = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\sqrt{\varrho_b}\, Y_{b,t} + \sqrt{1-\varrho_b}\, \epsilon_{i,t} \leq \Phi^{-1}(p_b)}$ 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

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

hence the squared default rate for infinitely many obligors satisfies

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

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 large. The additional adjustment brings the estimates closer to those in the limiting case of an infinite number of obligors. Note that we need a sufficiently large T (80 in Figure 3) for the additional adjustment to work well. Moreover, this adjustment applies only to the latent asset return correlation within a bucket, but not across buckets.

The difference between original value $\frac{1}{T}\sum_{t=1}^{T} x_t^2$ of the right-hand side of (14) and its adjusted value $\frac{1}{T}\sum_{t=1}^{T}(x_t^2 - x_t/N)$ is inversely proportional to the bucket size N. Since the bivariate normal cumulative distribution function Φ_2 and its inverse with respect to the correlation are continuous, the convergence of the adjustment implies that the adjusted estimator converges to the original estimator when N tends to infinity. While the convergence rate of the adjustment itself is $1/N$, the convergence speed of the estimator depends on the relation between $\frac{1}{T}\sum_{t=1}^{T} x_t^2$ and $\frac{1}{T}\sum_{t=1}^{T} x_t$: when default rates are high, making the relative difference between x_t^2 and x_t small, the estimator converges faster than when default rates are low.

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 $\tilde{x}_1, \dots, \tilde{x}_T$ 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.

This section and Section 4.1 demonstrate how method-of-moment estimators can be adjusted for idiosyncratic risk and autocorrelation in the underlying data. In contrast, such adjustments are not easily doable for MLE, which is an alternative to method-of-moment estimators. We give a detailed comparison with MLE in the next section when the idiosyncratic risk of obligors is negligible on the bucket level. We already note here that an MLE with AR(1) specification is not computationally feasible by standard methods when idiosyncratic risk is present. This is the reason why we did not include such a comparison in the current section.

4.3 A performance comparison with an MLE

Let us briefly recall how an MLE can be used in the current setting. 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

$$L(\varrho_b) = \int_{\mathbb{R}^T} p_b(y_t)^{D_t}\,(1-p_b(y_t))^{N_t-D_t}\,\mathrm{d}F(y_1,\dots,y_T) \qquad (16)$$

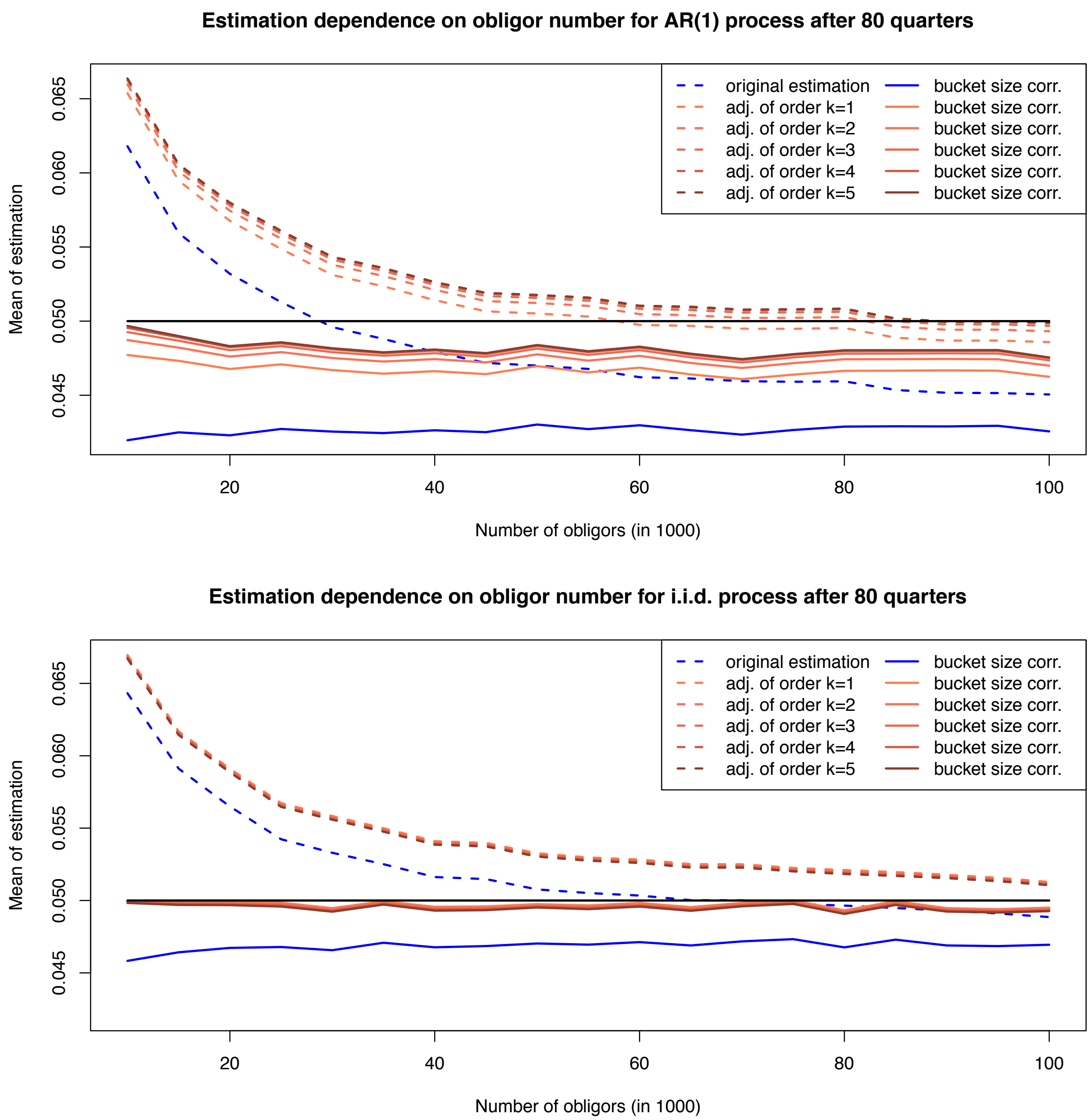

Figure 3: Influence of the idiosyncratic risk on the correlation estimations, as a function of the number of obligors in the bucket. For a finite number of obligors, idiosyncratic risk is present in the default rate time series, leading typically to higher correlation estimates (dashed curves, which 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 large.

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

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

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.,

$$\begin{aligned} X_t^{(k)} &= \sqrt{\varrho_b}\, Y_t + \sqrt{1-\varrho_b}\, \Xi_t^{(k)} \qquad k = 1, \ldots, N_t, \\ Y_t &= \rho\, Y_{t-1} + \sqrt{1-\rho^2}\, \Upsilon_t, \qquad Y_{t-1} \perp \Upsilon_t, \qquad t = 2, \ldots, T, \end{aligned}$$

the likelihood function takes the form

$$L(\varrho_b, \rho, \mu) = \int_{\mathbb{R}^T} p(y_t)^{D_t} \left(1 - p(y_t)\right)^{N_t - D_t} \mathrm{d}\Phi_T((y_1, \ldots, y_T)'; \mu, \Sigma_T), \tag{17}$$

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

$$\Sigma_T = \begin{pmatrix} 1 & \rho & \cdots & \rho^T \\ \rho & 1 & \cdots & \rho^{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^T & \rho^{T-1} & \cdots & 1. \end{pmatrix} \tag{18}$$

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 and when the idiosyncratic risk is diversified away because of a large number of obligors in the rating bucket, the computation can be simplified, as mentioned in Section III of Miu and Ozdemir (2009). In this case, we can apply a transformation of probability density to (1) and $(Y_1, \ldots, Y_T) \sim \mathcal{N}(0, \Sigma_T)$ with Σ_T given in (18) to write the joint probability density function of $\big(p(Y_1), \ldots, p(Y_T)\big)$ as

$$\frac{\varphi_T\big(h(x_1), \ldots, h(x_T); 0, \Sigma_T\big)}{\frac{\varrho_b^{T/2}}{(1-\varrho_b)^{T/2}} \prod_{j=1}^{T} \varphi\Big(\frac{\Phi^{(-1)}(p_b) - \sqrt{\varrho_b}\, h(x_j)}{\sqrt{1-\varrho_b}}\Big)} \quad \text{with } h(x) = \frac{\Phi^{(-1)}(p_b) - \Phi^{(-1)}(x)\sqrt{1-\varrho_b}}{\sqrt{\varrho_b}}, \tag{19}$$

where $\varphi(x)$ is the standard normal probability density function and $\varphi_T(x_1, \ldots, x_T; 0, \Sigma_T)$ is the multivariate normal probability density function with mean 0 and covariance matrix Σ_T. Formula (19) yields the likelihood function as a product that can be used to numerically find the MLE. Note, however, that this needs the knowledge or assumption that the systematic returns follow an AR(1) process, in addition to the assumption that the idiosyncratic risk is diversified away on the bucket level. In general, MLE for dependent processes is computationally very demanding and can only be approximated by using sophisticated numerical techniques (see McNeil and Wendin (2007), and Wendin (2006)). 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 next give in Figure 4 a performance comparison of method of moments (MoMs) and MLE for an AR(1) process and i.i.d. observations. For both AR(1) and i.i.d. cases, we choose the MLE (16), assuming serial independence since we suppose that the underlying autocorrelation is not known. In the case of the AR(1) process, we additionally use an MLE based on the AR(1) specification from (19) although this estimator needs the knowledge or assumption that the systematic returns follow an AR(1) process, which is difficult to check in reality given the short time series and which is not used by the other estimators.

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 with i.i.d. specification (bias: 9.2%). The bias (6.8%) of the MLE with AR(1) specification is smaller than those of the original MoM estimator and the MLE with i.i.d. specification, but still much larger than that of the adjusted MoM estimator. In the upper panel of Figure 4, we see particularly that the MLE with i.i.d. specification has a higher risk of extremely underestimating the true correlation value, compared to the MoM estimators. Standard deviations across the estimators are similar, with original MoM and MLEs on roughly the same level, and the adjusted MoM on a slightly higher level.

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 MLEs when the dynamics of the underlying time series is not explicitly known.

4.4 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[8], we compute the Value-at-Risk (VaR)

[8] Paragraph 178 in Basel Committee on Banking Supervision. International convergence of capital measures and capital standards, 2006.

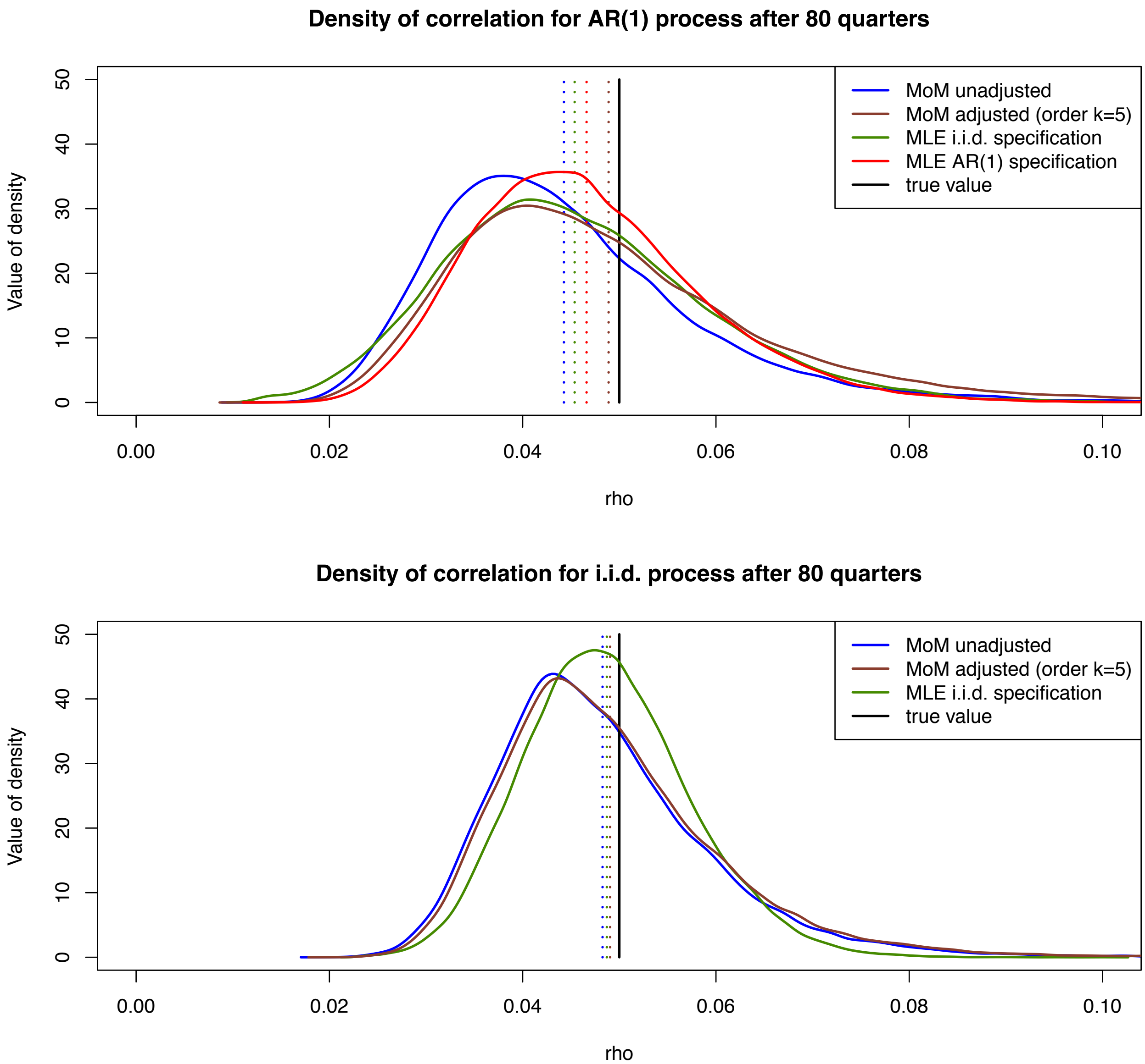

Figure 4: Comparison of distributions of the original method-of-moment (MoM) estimator, the 5th-order adjusted MoM estimator and MLEs with AR(1) and i.i.d. specifications. The underlying factors are simulated based on an AR(1) process with coefficient 0.7 (upper panel) and i.i.d. observations (lower panel). The solid curves are density estimations and the dotted lines give sample means. In the case of i.i.d. observations, the estimators perform similarly. For the AR(1) process, the bias of the MLE with i.i.d. specification (bias = 9.2%, std. dev. = 0.0135) is similar to that of the original MoM estimator (bias = 11.5%, std. dev. = 0.0135) while the bias of MLE with AR(1) specification (bias = 6.8%, std. dev. = 0.0118) is smaller, but still much larger 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 $\varrho_b = 0.05$, 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.

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 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.[9]

		true value	MoM unadjusted	MoM adjusted ($k = 5$)	MLE i.i.d. specific.	MLE AR(1) specific.
$\varrho_b = 0.05$	EL	3.99	3.99 (+0.05%)	3.99 (+0.09%)	4.00 (+0.10%)	3.99 (+0.07%)
	$\text{VaR}_{0.95}$	6.98	6.82 (−2.33%)	6.99 (+0.14%)	6.87 (−1.70%)	6.90 (−1.16%)
	$\text{VaR}_{0.99}$	9.26	8.94 (−3.44%)	9.31 (+0.54%)	9.03 (−2.47%)	9.11 (−1.59%)
	$\text{VaR}_{0.999}$	12.71	12.15 (−4.45%)	12.87 (+1.21%)	12.31 (−3.13%)	12.46 (−1.91%)
	$\text{VaR}_{0.9995}$	13.80	13.15 (−4.68%)	13.99 (+1.41%)	13.34 (−3.26%)	13.53 (−1.97%)
$\varrho_b = 0.10$	EL	3.99	4.00 (+0.33%)	4.01 (+0.49%)	4.00 (+0.14%)	4.00 (+0.25%)
	$\text{VaR}_{0.95}$	8.49	8.17 (−3.76%)	8.45 (−0.40%)	8.35 (−1.63%)	8.38 (−1.21%)
	$\text{VaR}_{0.99}$	12.75	11.98 (−6.04%)	12.69 (−0.46%)	12.40 (−2.74%)	12.51 (−1.90%)
	$\text{VaR}_{0.999}$	20.00	18.37 (−8.13%)	19.94 (−0.27%)	19.24 (−3.78%)	19.51 (−2.45%)
	$\text{VaR}_{0.9995}$	22.43	20.50 (−8.61%)	22.38 (−0.22%)	21.53 (−4.00%)	21.86 (−2.56%)

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_α numbers are closer to their true values for lower levels α and lower asset return correlation ϱ_b. Generally, our adjusted estimator (adjustment of order $k = 5$) leads to estimated VaR closest to the true values, better than the MLEs 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

[9] 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.

MoM, \$12.12m (−4.68%) for the MLE with i.i.d. specification, and \$12.27m (−3.49%) for the MLE with AR(1) specification.

Additionally to the through-the-cycle VaR estimates displayed in Table 1, our method provides an efficient means of computing point-in-time VaR. Figure 5 shows such point-in-time VaR estimates when the underlying asset returns are simulated by an AR(1) process over 80 quarters and the VaR is computed conditional on the last observation of the AR(1) process. We observe that the impact of our adjustment is pronounced for VaR at high levels and when current asset returns are low, corresponding to contraction periods in the credit cycle. The reason for this pronounced impact is that the autocorrelation, which is disregarded in the classical MoM estimator, strongly affects VaR at high levels and in downturn scenarios.

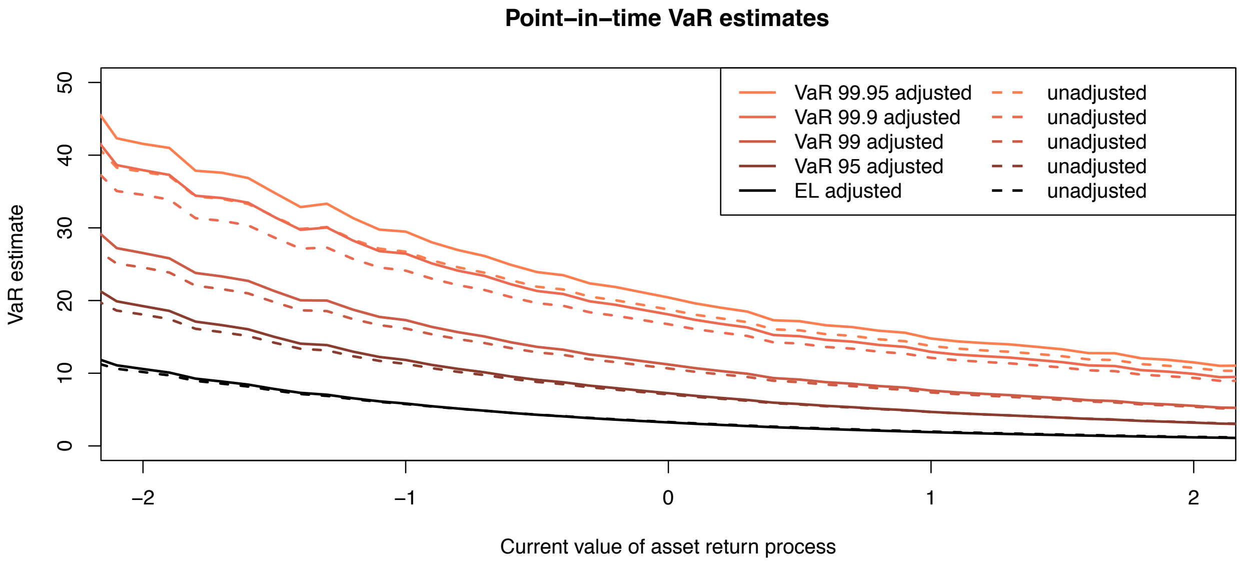

Figure 5: Point-in-time estimations of EL and VaR in an example of a diversified \$1 bn loan portfolio. EL and VaR at different levels (on vertical axis) over a 4-quarter horizon with 50,000 simulated paths are displayed conditional on different possibilities of the currently observed asset returns (on horizontal axis). The solid curves are done using our adjusted MoM estimator (adjustment of order $k = 5$) while the dashed curves are based on an unadjusted MoM estimator. The underlying factors are simulated based on an AR(1) process with coefficient 0.7, with asset return correlation $\varrho_b = 0.1$, and quarterly probability of default equal to 0.2%. Taking the averages of the EL and VaR estimates weighted by the distribution of asset returns results in the values in Table 1.

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 MLEs 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_α 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

$\varrho_b = 0.10$, $\mathrm{VaR}_{0.999}$ is underestimated by around 8% and 3% by unadjusted MoM and MLEs, respectively, while our adjusted MoM estimator provides essentially unbiased VaR numbers.

5 Conclusion

Starting with the observation that time series of default rates 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 a finite time series, when autocorrelation is present, the slower convergence leads to a larger bias of classical estimators. To ameliorate this shortcoming, we constructed an estimator that includes correction terms that take into account the autocorrelation and shortness of the observed time series. Applied to credit risk modelling, we found that our estimator eliminates much of the downward bias occurring in classical estimators of the latent asset return correlation. 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.

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

Altman, E., and Rijken, H. (2006). A point-in-time perspective on through-the-cycle ratings. *Financial Analysts Journal* **62**(1), 54–70

Bluhm, C., and Overbeck, L. (2003). Systematic risk in homogeneous credit portfolios. In: Bol, G., Nakhaeizadeh, G., Rachev, S., Ridder, T. and Vollmer, K.-H. (eds.): Credit Risk: Measurement, Evaluation and Management (Contributions to Economics), Springer

Bluhm, C., Overbeck, L., and Wagner, C. (2010). An Introduction to Credit Risk Modeling. Chapman & Hall/CRC, 2nd edition

Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *Journal of Time Series Analysis* **17**(3), 247–270, DOI:10.1111/j.1467-9892.1996.tb00275.x

Cantor, R., and Mann, C. (2006). Analyzing the tradeoff between ratings accuracy and stability. *Journal of Fixed Income*, September 2006, DOI:10.3905/jfi.2007.683318

Carey, M., and Hrycay, M. (2001). Parametrizing credit risk models with rating data. *Journal of Banking and Finance* **25**(1), 197–270, DOI:10.1016/S0378-4266(00)00124-2

Durrett, R. (2010). Probability: Theory and Examples. 4th edition, Cambridge University Press

Frey, R. and McNeil, A. J. (2003). Dependent defaults in models of portfolio credit risk. *Journal of Risk* **6**(1), 59–92, DOI:10.21314/JOR.2003.089

Gordy, M. (2000). A comparative anatomy of credit risk models. *Journal of Banking and Finance* **24**(1–2), 119–149, DOI:10.1016/S0378-4266(99)00054-0

Gordy, M., and Heitfield, E. (2010). Small sample estimation of models of portfolio credit risk. In: Kijima, M., Hara, C., Tanaka, K., and Muromachi, Y. (eds.): Recent Advances in Financial Engineering: The Proceedings of the KIER-TMU International Workshop on Financial Engineering 2009, World Scientific Publishing

Gordy, M., and Howells, B. (2006). Procyclicality in Basel II: Can we treat the disease without killing the patient? *Journal of Financial Intermediation* **15**(3), 395–417, DOI:10.1016/j.jfi.2005.12.002

Götze, F., and Künsch, H.R. (1996). Second order correctness of the blockwise bootstrap for stationary observations. *The Annals of Statistics* **24**(5), 1914–1933

Greene, W.H. (2008). Econometric Analysis. 6th edition, Pearson Prentice Hall

Hamilton, J. (1994). Time Series Analysis, Princeton University Press

Heitfield, E. (2004). Rating system dynamics and bank-reported default probabilities under the New Basel Capital Accord. Working paper, Federal Reserve Board

Löffler, G. (2004). An anatomy of rating through the cycle. *Journal of Banking and Finance* **28**(3), 695–720, DOI:10.1016/S0378-4266(03)00041-4

Löffler, G. (2005). Avoiding the rating bounce: why rating agencies are slow to react to new information. *Journal of Economic Behavior and Organization* **56**(3), 365–381, DOI:10.1016/j.jebo.2003.09.015

McNeil, A., Frey, R., and Embrechts, P. (2015). Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press

McNeil, A., and Wendin, J. (2006). Dependent credit migrations. *Journal of Credit Risk* **2**(2), 87–114, DOI:10.21314/JCR.2006.043

McNeil, A., and Wendin, J. (2007). Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance* **14**(2), 131–149, DOI:10.1016/j.jempfin.2006.05.002

Merton, R. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* **29**(2), 449–470, DOI:10.2307/2978814

Miu, P., and Ozdemir, B. (2009). Stress-testing probability of default and migration rate with respect to Basel II requirements. *Journal of Risk Model Validation* **3**(4), 1–36, DOI:10.21314/JRMV.2009.048

Nagpal, K., and Bahar, R. (2001). Measuring default correlation. *Risk* **14**(3), 129–132

Shumway, R.H., and Stoffer, D.S. (2017). Time Series Analysis and Its Applications: With R Example. 4th edition, Springer

Wendin, J. (2006). Bayesian Methods in Portfolio Credit Risk Management. *PhD Thesis, ETH No. 16481*, DOI:10.3929/ethz-a-005182512