

MODELING METHODOLOGY

Authors

Jimmy Huang
Libor Pospisil

Acknowledgements

We would like to thank Amnon Levy and Christopher Crossen for their comments.

Contact Us

Americas
+1.212.553.1658
clientservices@moodys.com

Europe
+44.20.7772.5454
clientservices.emea@moodys.com

Asia (Excluding Japan)
+85 2 2916 1121
clientservices.asia@moodys.com

Japan
+81 3 5408 4100
clientservices.japan@moodys.com

Uncertainty in Asset Correlation Estimates and Its Impact on Credit Portfolio Risk Measures

Abstract

Credit portfolio models rely on estimated and calibrated parameters, such as default and rating migration probabilities, recovery rates, and asset correlations. Users of these models must understand how various errors in the parameter estimates impact model outputs, for example Unexpected Loss (UL) or Economic Capital (EC).

Asset correlations estimated using asset return time series are subject to inherent uncertainty — statistical errors — arising due to a limited length of the time series. The main question this paper addresses is how these errors translate into statistical errors in the estimated UL and EC.

We illustrate several properties of the errors using an analytical method. As expected, longer time series lead to lower errors in UL and EC. Increasing the number of exposures in a portfolio, however, can reduce the errors in UL and EC only to a certain degree.

We also conduct simulation studies quantifying the impact of statistical errors in correlation estimates on UL and EC for realistic portfolios. Our studies focus on three types of correlation models — a pairwise structure free model, a single-factor model, and a multi-factor model similar to Moody's Analytics Global Correlation Corporate Model (GCorr™ Corporate).

Table of Contents

1. Introduction	3
2. Link Between Uncertainty in Correlations and in Portfolio Risk Measures	5
2.1 Sample Pairwise Correlation and Joint Default Probability	5
2.2 Uncertainty in Portfolio Risk Measures as a Function of Uncertainty in Average Correlations	7
3. Properties of Uncertainty in Correlations	11
3.1 Uncertainty in Average Correlation	11
3.2 Factor Model	14
4. Impact of Uncertainty in Correlations on Risk Measure Estimates	18
4.1 Portfolio-Level Risk Measures	18
4.2 Decomposition of Uncertainty in a Multi-Factor Model	22
4.3 Portfolio-Referent Exposure-Level Risk Measures	23
5. Summary	25
Appendix A Uncertainty in a Single-Factor Model	26
Uncertainty in a Pairwise Correlation	26
Covariance of Two Pairwise Correlation Estimates	26
Appendix B Uncertainty in a Risk Contribution	28
References	30

1. Introduction

Models that quantify a credit portfolio's risk use various parameters as inputs. Examples include default and migration probabilities of the firms within the portfolio, recovery rates, asset correlations, or other measures of correlations. These parameters must be estimated using historical data or determined in another way if sufficient data is not available, such as using expert judgment or utilizing parameter values of firms with similar credit risk characteristics. Estimated parameters may differ from the true values of the parameters, either because of bias or due to uncertainty (also referred to as statistical error or sampling error) inherently present in any estimate based on a sample from a population.¹

This paper discusses statistical errors in estimated asset correlations² and their impact on a credit portfolio's various risk measures. These risk measures, namely Unexpected Loss (UL) and Economic Capital (EC), are a function of asset correlations among the firms for which the credit portfolio has exposure. Thus, uncertainty in asset correlations used as inputs for calculations translates into uncertainty in the resulting value of a risk measure. In addition to portfolio-level risk measures, we consider uncertainty in Risk Contribution (RC) — portfolio-referent capital allocated to individual instruments.

We can use various types of data sources to estimate asset correlations — using firm-level time series of asset returns or using default rate time series at a level of homogeneous pools. This paper assumes that firm-level time series of asset returns (or other time series approximating changes in credit qualities) are available to estimate correlations.³

While the properties of stand-alone estimated asset correlations are well-known, we emphasize the need to understand them within a portfolio context. We focus on three approaches to estimating correlations for a portfolio — structure-free empirical pairwise correlations, a single-factor model, and a multi-factor model similar to Moody's Analytics Global Correlation Corporate Model (GCorr Corporate). Our first objective is to understand properties of statistical errors in estimated correlations and their impacts on risk measures using analytical methods. Specifically, we examine how the errors depend on portfolio size, correlation levels, method of estimation, sample size, and other characteristics. Second, we conduct simulation studies that illustrate numerically the magnitude of uncertainty in EC due to the statistical errors in estimated correlations.

Portfolio credit risk model developers and users should understand the impact of errors in input parameters because such analysis is an integral part of model risk assessment process. The recent financial crisis shows how improper use of financial models, including unreasonable assumptions and incorrect parameter values, leads to substantial losses for financial institutions. Regulators have published several documents on the topic of model risk in the past. One of the most important of these documents is a regulatory guideline on model risk management within financial institutions issued by the Federal Reserve System and the OCC, known as SR 11-7.⁴ The guideline discusses uncertainty regarding model inputs as one of the primary components of model risk that financial institutions should be attentive to.

Existing academic papers relevant to our research can be divided into two categories. The first includes papers focused directly on credit portfolios. Löffler (2003) employs Monte Carlo simulations within the Credit Metrics modeling framework to quantify statistical errors in the credit Value-at-Risk (VaR) measure. VaR is determined using three sets of estimated parameters — default probabilities, recovery rates, and asset correlations. According to their results, statistical errors in default probabilities tend to impact VaR most. However, if the target probability becomes low and the default probabilities of firms high, the effect of statistical errors in asset correlations grows in importance. Tarashev and Zhu (2007) examine how various assumptions and inputs of the Asymptotic Single-Risk Factor (ASRF) model impact a realistic portfolio's Economic Capital. The authors estimate structure-free, pairwise asset correlations and utilize them to calibrate a single-factor model, which is, in turn, used for Economic Capital calculation. Results suggest that the statistical errors in correlations lead to a relatively large uncertainty in the Economic Capital. Increasing the size of the portfolio does not mitigate this effect. While long time series used for estimation lead to a lower uncertainty in Economic Capital, the uncertainty remains notable even for realistically long time series.

This paper focuses on asset correlations estimated from time series of firm-level asset returns. Other estimation methods — within the context of credit risk — include using the default rate time series of a homogenous pool. Gordy and Heitfield (2002), for example, estimate precision of correlations estimated in this way.

¹ The terms "uncertainty," "statistical error," and "sampling error" in an estimated parameter refer to the same standard statistical concept in this document — deviation of the estimated parameter from its expected value due to the fact that only a sample of data is used for estimation. If no bias is present, the expected value is equal to the true value of the parameter.

² Throughout this document, we use the terms "asset correlation" and "correlation" interchangeably.

³ It is possible to interpret time series of factors implied by default rates as systematic changes in credit qualities and apply the results from this paper. However, implying a factor time series or a within-pool correlation from a default rates leads to further uncertainties, due to possible non-homogeneity of the pool or its small size, and we do not address these questions in the paper.

⁴ Board of Governors of the Federal Reserve System (2011), "Guidance on Model risk Management."

The second category of relevant papers does not consider credit portfolios as in this paper, but focuses on stock portfolios instead. Various stock investment strategies, such as Markowitz mean-variance optimization, depend on estimated input parameters, including stock return correlations, equivalent to asset return correlations from a statistical error perspective. Chopra and Ziemba (1993) compare how statistical errors in expected stock returns, return variances, and return covariances affect choosing the optimal stock portfolio. The authors conclude that errors in sample return covariances distort portfolio selection less than errors in expected returns. Broadie (1993) uses Monte Carlo simulations to illustrate how statistical errors in estimated parameters lead to large deviations of the estimated efficient frontier from the true efficient frontier. The paper also discusses the tradeoff between using shorter and longer time series for parameter estimation. While the shorter time series may provide more current parameter values, they naturally lead to larger statistical errors.

We organize the remainder of the paper as follows:

- » Section 2 motivates the discussion and creates a framework for analyzing the impact of uncertainty in correlations on the uncertainty in credit portfolio risk measures.
- » Section 3 relies on the framework from Section 2 to study analytically how uncertainty in credit portfolio risk measures as a result of correlation uncertainty depends on various data and portfolio characteristics — number of observations available for estimation, portfolio size, correlation level, and credit risk level.
- » Section 4 presents simulation exercises illustrating — based on realistic portfolios — the extent to which correlation uncertainty affects credit portfolio risk measures.
- » Section 5 concludes.
- » Appendix A contains derivation of formulas for uncertainty in correlation estimates in a single-factor model.
- » Appendix B provides intuition for patterns in uncertainty in estimates of the Risk Contribution.

2. Link Between Uncertainty in Correlations and in Portfolio Risk Measures

This section discusses basic properties of correlation estimates and how their uncertainty plays a role in estimating portfolio risk measures.⁵

2.1 Sample Pairwise Correlation and Joint Default Probability

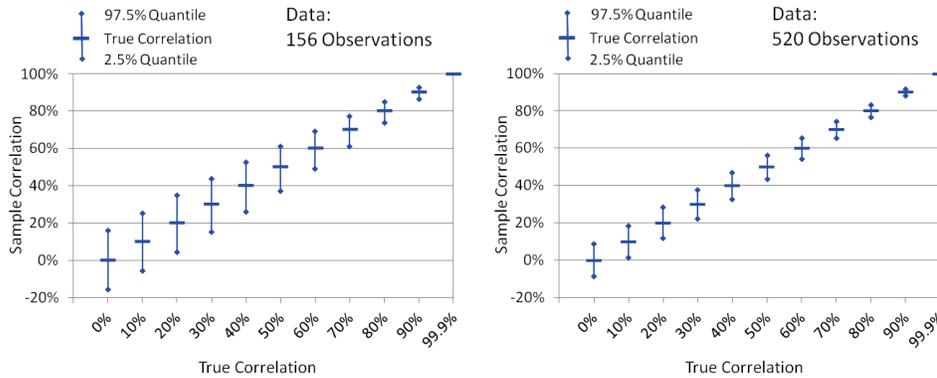
Consider two time series of asset returns of length T (where T can represent for example the number of weekly observations): r_{1t} and r_{2t} , $t = 1, \dots, T$. We assume these series are serially uncorrelated and have the same joint normal distribution (r_{1t}, r_{2t}) for all t . The sample correlation⁶ of these series, $\hat{\rho}$, serves as an estimator of the correlation parameter associated with the joint distribution: $\rho = \text{corr}(r_{1t}, r_{2t})$.

The asymptotic properties of the sample correlation⁷ can be described through its so-called Fisher's z-transformation $\frac{1}{2} \log \left(\frac{1+\hat{\rho}}{1-\hat{\rho}} \right)$. If the length of the time series is sufficiently large, then according to the well-known result, the transformed sample correlation has a normal distribution with variance independent of the correlation level:

$$\frac{1}{2} \log \left(\frac{1+\hat{\rho}}{1-\hat{\rho}} \right) \sim N \left(\frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right); \frac{1}{T-3} \right) \tag{1}$$

The distribution in Equation (1) gives us a way to plot the distribution of the sample correlation for various levels of correlations ρ and numbers of observations T . Figure 1 shows several examples of the distribution.

Figure 1 Sampling distribution of empirical pairwise correlations based on Fisher's z-transformation.



⁵ Throughout the rest of this paper, the term "portfolio" always refers to "credit portfolio."

⁶ We use the standard definition of sample correlation: $\hat{\rho} = \frac{T \sum_{t=1}^T r_{1t} r_{2t} - \sum_{t=1}^T r_{1t} \sum_{t=1}^T r_{2t}}{\sqrt{T \sum_{t=1}^T r_{1t}^2 - (\sum_{t=1}^T r_{1t})^2} \sqrt{T \sum_{t=1}^T r_{2t}^2 - (\sum_{t=1}^T r_{2t})^2}}$

⁷ We use the terms "sample correlation," "empirical correlation," and "historical correlation" interchangeably.

2.5% and 95.5% Quantiles of the Sample Correlation Distribution

True Correlation ρ	Data: 156 Observations		Data: 520 Observations	
	2.5% Quantile	97.5% Quantile	2.5% Quantile	97.5% Quantile
0%	-15.7%	15.7%	-8.6%	8.6%
1%	-14.7%	16.7%	-7.6%	9.6%
5%	-10.8%	20.6%	-3.6%	13.5%
10%	-5.8%	25.3%	1.4%	18.4%
20%	4.4%	34.6%	11.6%	28.1%
30%	15.0%	43.7%	22.0%	37.6%
40%	25.9%	52.4%	32.5%	47.0%
50%	37.2%	60.9%	43.3%	56.2%
60%	48.9%	69.2%	54.2%	65.2%
70%	61.0%	77.2%	65.3%	74.1%
80%	73.5%	85.0%	76.7%	82.9%
90%	86.5%	92.6%	88.2%	91.5%
99.9%	99.9%	99.9%	99.9%	99.9%
100%	100%	100%	100%	100%

One pattern that stands out in Figure 1 is the high magnitude of uncertainty in sample correlation for realistic sample sizes. For example $T = 156$ represents three years of weekly time series, which can be interpreted as data used for the estimate of a point-in-time correlation. However, if the true correlation is 20%, the sample correlation can range over a fairly wide interval from 4.4%–34.6%. Introducing longer time series, for example ten years of weekly data or $T = 520$ with the intention of estimating through-the-cycle parameters, leads to a narrowing of the typical interval for sample correlations, but it still spans from 11.6%–28.1%. Given this amount of statistical error in correlation estimates, we want to examine two questions: how does this uncertainty impact credit risk measurement? And does the statistical error change if we move away from a single pair of firms to the credit risk analysis for a large portfolio?

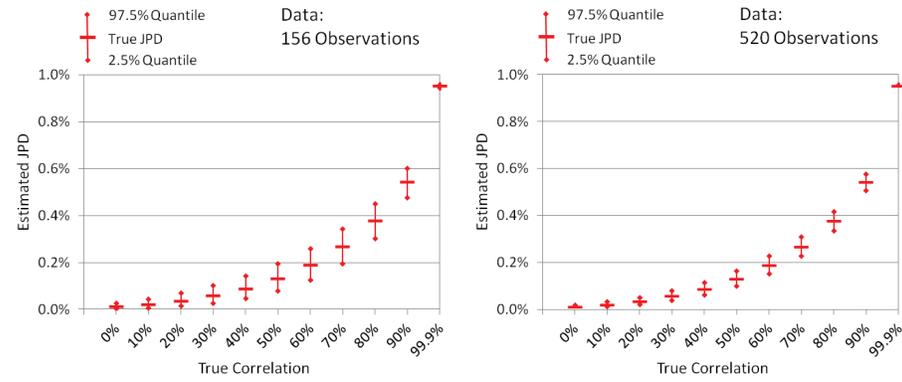
Asset return correlations matter in a credit risk context for several reasons. Defaults, rating downgrades, and credit spread movements cause credit losses, and it is important to understand to what degree these events occur for multiple firms at the same time, thus magnifying portfolio losses. The concurrence of defaults (and similarly other events) can be measured using asset correlations. For example, if two firms have known default probabilities PD_1 and PD_2 , and if they default when $r_1 < N^{-1}(PD_1)$ or $r_2 < N^{-1}(PD_2)$, then the estimated joint default probability (also called joint expected default frequency), $J\widehat{PD}$, can be expressed as a function of the empirical correlation:

$$J\widehat{PD} = N_2(N^{-1}(PD_1), N^{-1}(PD_2); \hat{\rho})$$

N_2 denotes two-dimensional normal cumulative distribution function with standard normal marginal distributions and a correlation parameter (2)

Figure 2 shows the distribution of the estimated joint probability for various correlation levels. For a true correlation level of 20% — realistic correlation for corporate firms — and marginal PDs of 1%, and assuming three years of weekly data for correlation estimation, the distribution of the estimated joint default probability ranges from 14bps–69bps, with the true value being 34bps. In other words, the probability estimate could be, with reasonable likelihood, as low as half of the true probability or as high as twice the true probability. While correlations are well established parameters, it is only when we move to the parameters with more straightforward real-world interpretation in credit risk context — such as joint expected default frequency — that the implications of statistical errors become clearer.

Figure 2 Distribution of estimated joint default probability based on sample correlation. Assumption: $PD_1 = 1\%$ and $PD_2 = 1\%$.



2.5% and 95.5% Quantiles of the Estimated JPD Distribution

True Correlation ρ	True JPD	Data: 156 Observations		Data: 520 Observations	
		2.5% Quantile	97.5% Quantile	2.5% Quantile	97.5% Quantile
0%	0.010%	0.003%	0.027%	0.005%	0.018%
1%	0.011%	0.003%	0.028%	0.006%	0.019%
5%	0.014%	0.004%	0.035%	0.008%	0.024%
10%	0.019%	0.006%	0.044%	0.011%	0.031%
20%	0.034%	0.014%	0.069%	0.021%	0.051%
30%	0.056%	0.026%	0.101%	0.038%	0.078%
40%	0.087%	0.046%	0.142%	0.062%	0.115%
50%	0.129%	0.077%	0.194%	0.099%	0.163%
60%	0.188%	0.124%	0.259%	0.152%	0.226%
70%	0.267%	0.194%	0.342%	0.227%	0.308%
80%	0.377%	0.301%	0.450%	0.336%	0.417%
90%	0.542%	0.475%	0.602%	0.506%	0.576%
99.9%	0.952%	0.944%	0.959%	0.948%	0.956%
100%	1.000%	1.000%	1.000%	1.000%	1.000%

Figure 2 also illustrates the highly non-linear nature of the problem. The absolute and relative magnitudes of uncertainty in the joint default probability depend—through function N_2 in Equation (2)—on the magnitude of uncertainty in correlation, level of correlation, and the PD parameters. Equation (2) thus adds another layer of complexity compared to analyzing stock portfolios, where losses are directly linear functions of returns r .

Having demonstrated the impact of uncertainty in the correlation of a single pair of firms on credit risk measurement, we consider a portfolio containing multiple firms next.

2.2 Uncertainty in Portfolio Risk Measures as a Function of Uncertainty in Average Correlations

Assume a portfolio of credit exposures to K corporates. In order to discuss portfolio risk measures, we leverage a credit portfolio model along the lines of the framework developed by Moody's Analytics, described by Levy (2008). The framework begins by defining asset returns r_k ($k = 1, \dots, K$) as normal random variables representing changes in the credit qualities of firms between the analysis date and a future horizon. Asset correlations describe the dependence among asset returns of various firms. A specific value a firm's asset return implies the firm's asset value and, therefore, credit quality at the horizon, which, in turn, can be associated with a specific value of a credit instrument issued by the firm. Given a credit portfolio characteristics (types of exposures, exposure amounts, maturities, cash-flow structures, etc.) and specific input parameter values (asset correlations, default probabilities, credit migration matrix, loss given defaults, etc.), the modeling framework uses Monte-Carlo simulation of asset returns to estimate value/loss distribution of the credit portfolio at the horizon (e.g. one year). This distribution can be used to determine risk measures, such as UL or EC . Moreover, calculation outputs also include individual instrument's RC in the portfolios.

We denote a portfolio risk measure as RM (where $RM = UL, EC, \text{ or } ES$). From a theoretical perspective, the risk measure can be interpreted as a function f of portfolio characteristics and input parameters. As an example, Equation (3) shows RM as a function

of pairwise asset correlations $\rho_{kl} = \text{corr}(r_k, r_l)$ across all pairs of K borrowers in the portfolio⁸ and *Other Inputs*, which can include default probabilities, credit migration matrix, and further parameters.

$$RM = f(\{\rho_{kl}\}_{1 \leq k < l \leq K}; \text{Other Inputs}) \quad (3)$$

The correlation parameters in Equation (3) can be either specified as a structure free pairwise correlation matrix, or it can have factor structure imposed upon them — a single-factor model or a multi-factor model (along the lines of the GCorr Corporate model developed by Moody's Analytics and described in paper by Huang, et al. (2012a)), in which case, the information regarding pairwise correlations can be summarized using fewer parameters:

$$\begin{aligned} \text{Single-factor model} - \rho_{kl} &= \rho_{k\phi}\rho_{l\phi}, \text{ where } \rho_{k\phi} = \text{corr}(r_k, \phi) \text{ and } \phi \text{ is the systematic factor}^9 \\ \text{Multi-factor model} - \rho_{kl} &= \rho_{k\phi}\rho_{l\phi}\text{corr}(\phi_k, \phi_l), \text{ where } \rho_{k\phi} = \text{corr}(r_k, \phi_k) \text{ and } \phi_k \text{ is the systematic factor} \\ &\text{driving the asset return } r_k. \text{ }^{10} \end{aligned} \quad (4)$$

In practice, most input parameters are estimated from data and possibly adjusted by expert judgment. Moreover, the portfolio value distribution can be determined exactly only under very special circumstances (such as default/no-default valuation within a single-factor model); typically, a Monte-Carlo simulation is used. Incorporating all the uncertainties into Equation (3) leads to the following expression:

$$\widehat{RM} = \widehat{f}(\{\widehat{\rho}_{kl}\}_{1 \leq k < l \leq K}; \text{Other Inputs}) \quad (5)$$

Terms $\widehat{\rho}_{kl}$ and *Other Inputs* represent, respectively, asset correlation estimates and estimates of other input parameters, such as default probabilities. Finally, \widehat{f} approximates the portfolio risk measure through, for example, Monte-Carlo simulation (if an analytical formula is used for risk measure calculation, then $\widehat{f} = f$).

In further discussion, we consider RM from Equation (3) the true value of the risk measure — we assume the model correctly specifies the shape of asset return distributions (for example, normal distributions), the dependence structure (for example a Gaussian multi-factor model), and further, the model correctly values instruments in the portfolio. Given these assumptions, the statistical error in the estimated risk measure, $\text{Err}_{RM} = \widehat{RM} - RM$, can come from three broad sources and their possible interactions:

- » Simulation noise or approximation error due to the fact that f is not known and must be estimated via simulation or a numerical approximation procedure.
- » Biases in parameter estimates. The input parameters, such as correlations and default probabilities, are estimated from data and/or by applying expert judgment. The estimation process can lead to biased estimates for various reasons — either the estimation method itself may be biased,¹¹ the data used for estimation is unsuitable (for example, choosing too long a time period despite the fact that the parameter of interest might have changed during the period¹²), or incorrect expert judgment.
- » Statistical errors in parameters. Any estimation method based on a finite data sample, which does not represent the entire population, leads to parameter estimates subject to statistical errors or deviations from their expected (or population) values. All of the credit portfolio model's input parameters, estimated based on data, such as correlations and default probabilities, contribute to this error.

This paper focuses solely on the statistical errors in estimated correlations and their impact on portfolio risk measures. We assume that all other inputs, such as default probabilities, are given. In the analytical section of this paper, we assume that function f does not need approximating. However, when conducting simulation exercises for a multi-factor model, we must relax this assumption, as there is no analytical expression for f in that situation.

Considering the case where only uncertainty in correlations leads to uncertainty in portfolio risk measure, we have

⁸ Parameter ρ_{kl} is the correlation between asset returns of borrowers k and l , while $\{\rho_{kl}\}_{1 \leq k < l \leq K}$ denotes the vector of all unique pairwise correlations across borrowers $1, \dots, K$. The number of such correlations is $K(K-1)/2$.

⁹ This correlation is related to the so called RSQ value of the firm: $\rho_{k\phi} = \sqrt{RSQ_k}$

¹⁰ Similar, to the single factor case, this correlation is related to the so called RSQ value of the firm: $\rho_{k\phi} = \sqrt{RSQ_k}$.

¹¹ As an example, see the paper by Huang, et al. (2012b).

¹² There is a natural trade-off between using a longer time period of data (more observations, possibly non-stationary parameters) versus a shorter time period (fewer observations, stationary parameters). For a discussion of this topic within the context of estimating means, variances, and correlations of stock returns, see Broadie (1993).

$$\widehat{RM}_\rho = f(\{\widehat{\rho}_{kl}\}_{1 \leq k < l \leq K}; \text{Other Inputs}) \quad (6)$$

We further assume that the correlations are estimated without a bias, or at least a substantial bias, so that the main source of difference between the estimated value $\widehat{\rho}_{kl}$ and the true value ρ_{kl} can be attributed to the statistical error: $Err_{\rho_{kl}} = \widehat{\rho}_{kl} - \rho_{kl}$.¹³

We are interested in understanding the magnitude of $Err_{RM\rho} = \widehat{RM}_\rho - RM$, which we can measure by $std(Err_{RM\rho}) = std(\widehat{RM}_\rho)$. One of the main points of this paper is that this error does not depend only on the levels of the various input parameters and errors in pairwise correlation estimates $Err_{\rho_{kl}}$, but, crucially, also on the interactions of these pairwise correlation errors, which we illustrate in further discussion.

The pairwise correlation estimates are based on T observations of asset returns associated with firms $1, \dots, K$. For sufficiently large T , the joint distribution of the correlation estimates can be approximated by a normal distribution thanks to the Central Limit Theorem:

$$\sqrt{T}(\{\widehat{\rho}_{kl}\}_{1 \leq k < l \leq K} - \{\rho_{kl}\}_{1 \leq k < l \leq K}) \sim N(0; \Sigma), \text{ where } \Sigma = \{\sigma_{kl,ij}\}_{\substack{1 \leq k < l \leq K \\ 1 \leq i < j \leq K}} \quad (7)$$

It is worth highlighting that in the case of a single-pairwise correlation, the statement in Equation (7) is equivalent to the Fisher's z-transformation result in Equation (1). The Fisher's z-transformation is often used to speed up the convergence to normal distribution, because Equation (7) may require many observations (a large T) in order to hold, depending on the properties of the underlying asset return data. Another reason to use Fisher's z-transformation is to stabilize its variance of correlation estimates,¹⁴ which matters especially for the extreme values of correlations close to 1 or -1.

Note, the covariance matrix Σ in Equation (7) has dimension $K(K-1)/2 \times K(K-1)/2$. The formula implies that the asymptotic covariance between two correlation estimates is¹⁵ $cov_T(\widehat{\rho}_{kl}, \widehat{\rho}_{ij}) = \sigma_{kl,ij}/T$.

If f is continuously differentiable with respect to the correlation parameters ρ_{kl} ¹⁶, and meets several other assumptions, we can apply the so called delta method¹⁷ to deduce asymptotic distribution of the error in risk measure, $Err_{RM\rho}$. For a sufficiently large number of observations T , the delta method implies an approximate distribution of the error in the risk measure estimate:

$$\begin{aligned} \sqrt{T}(\widehat{RM}_\rho - RM) &\sim N(0; \sigma_{RM}^2) \\ \sigma_{RM}^2 &= \sum_{1 \leq k < l \leq K} \sum_{1 \leq i < j \leq K} \frac{\partial f}{\partial \rho_{kl}} \frac{\partial f}{\partial \rho_{ij}} \sigma_{kl,ij} \end{aligned} \quad (8)$$

The covariance terms $\sigma_{kl,ij}$ in Equation (8) come from Equation (7) and the partial derivative terms $\frac{\partial f}{\partial \rho_{kl}}$ are evaluated at the true values of asset correlations $\{\rho_{kl}\}_{1 \leq k < l \leq K}$.

If we measure uncertainty in estimates using standard deviation, Equation (9) implies the relationship between uncertainty in portfolio risk measure and the correlation estimates:

$$std_T(\widehat{RM}_\rho) = \sqrt{\frac{1}{T} \cdot \sum_{1 \leq k < l \leq K} \sum_{1 \leq i < j \leq K} \frac{\partial f}{\partial \rho_{kl}} \frac{\partial f}{\partial \rho_{ij}} \sigma_{kl,ij}} = \left| \sum_{1 \leq k < l \leq K} \frac{\partial f}{\partial \rho_{kl}} \right| \cdot std_T(\widehat{\rho}_{Avg}) \quad (9)^{18}$$

where the average correlation estimate is defined as

$$\widehat{\rho}_{Avg} = \sum_{1 \leq k < l \leq K} w_{kl} \cdot \widehat{\rho}_{kl}, \quad w_{kl} = \frac{\partial f / \partial \rho_{kl}}{\sum_{1 \leq k < l \leq K} \partial f / \partial \rho_{kl}}$$

¹³ If the estimator is consistent, the bias impact becomes small, as the number of observations increase.

¹⁴ Note, while the variance in Equation (1) does not depend on the correlation level, it does in Equation (7), through the term $\sigma_{kl,kl}$.

¹⁵ We use subscript T for terms where we want to stress dependence on the number of observations.

¹⁶ To illustrate that this assumption underpinning the delta method is realistic, we can point to the single factor case in Equation (24), Section 4, in which the capital is a continuously differentiable function of the correlation parameters.

¹⁷ The delta method is a standard statistical technique, described for example in Section 3 of van der Vaart (2000).

¹⁸ We note that if f is an increasing function of correlations, then the absolute value in the first equation is redundant. This would be the case of capital, for example, on an accrual loan portfolio.

and its standard deviation is

$$std_T(\widehat{\rho}_{Avg}) = \sqrt{\frac{1}{T} \cdot \sum_{1 \leq k < l \leq K} \sum_{1 \leq i < j \leq K} w_{kl} w_{ij} \sigma_{kl,ij}}$$

Equation (9) provides a framework to study how uncertainty in correlation estimates impacts uncertainty in portfolio risk measures. It includes the interactions of the individual correlation estimates through terms $\sigma_{kl,ij}$ and then conveniently summarizes the uncertainties of all correlation estimates into a single quantity — $std_T(\widehat{\rho}_{Avg})$ — the quantity that ultimately matters within the portfolio context.

Section 3 discussed properties of the uncertainty in the average correlation. The following section presents examples of correlation models and risk measures, where f is smooth and the terms $\frac{\partial f}{\partial \rho_{kl}}$ can be expressed analytically. These include *EC* in a single factor asymptotic model or *UL*. In some cases, such as *EC* for a small portfolio, f may not be a smooth function of correlations, but, even then, the general idea behind Equation (9) provides a useful framework for thinking about link between uncertainty in correlations and uncertainty in portfolio risk measures.

3. Properties of Uncertainty in Correlations

This section focuses on understanding properties of the uncertainty in average correlation estimator introduced in Equation (9), in the context of both a structure-free pairwise correlation case, as well as a single-factor correlation model.

3.1 Uncertainty in Average Correlation

The question of interest is how the uncertainty of average correlation estimate from Equation (9), $std_T(\widehat{\rho}_{Avg})$, depends on the number of observations T , size of the portfolio K , and the correlation levels $\{\rho_{kl}\}_{1 \leq k < l \leq K}$.

Dependence of the uncertainty on number of observations is fairly straightforward—the standard deviation of average correlation converges to zero at rate $1/\sqrt{T}$ as T becomes large:

$$std_T(\widehat{\rho}_{Avg}) \xrightarrow[1/\sqrt{T}]{} 0 \text{ as } T \rightarrow \infty \quad (10)^{19}$$

In order to express dependence on portfolio size, we must decompose the standard deviation $std_T(\widehat{\rho}_{Avg})$ in more detail. Namely, we consider three cases of covariances of correlation estimates $\sigma_{kl,ij}$: the variance of a single pairwise correlation ($(k, l) = (i, j)$), the case of two pairwise correlations where one asset return is common to both pairs ($(i, j) \neq (k, l)$ and $\{i, j\} \cap \{k, l\} \neq \emptyset$), and finally, the case of two pairwise correlations where the two pairs contain completely different asset returns ($(i, j) \neq (k, l)$ and $\{i, j\} \cap \{k, l\} = \emptyset$).

$$\begin{aligned} std_T(\widehat{\rho}_{Avg}) &= \sqrt{\frac{1}{T} \cdot \sum_{1 \leq k < l \leq K} \sum_{1 \leq i < j \leq K} w_{kl} w_{ij} \sigma_{kl,ij}} = \\ &= \sqrt{\frac{1}{T} \cdot \left[\sum_{1 \leq k < l \leq K} w_{kl}^2 \sigma_{kl,kl} + \sum_{1 \leq k < l \leq K} \sum_{\substack{1 \leq i < j \leq K \\ (i,j) \neq (k,l) \\ \{i,j\} \cap \{k,l\} \neq \emptyset}} w_{kl} w_{ij} \sigma_{kl,ij} + \sum_{1 \leq k < l \leq K} \sum_{\substack{1 \leq i < j \leq K \\ (i,j) \neq (k,l) \\ \{i,j\} \cap \{k,l\} = \emptyset}} w_{kl} w_{ij} \sigma_{kl,ij} \right]} \end{aligned} \quad (11)$$

A well-known result, presented by, for example, Steiger (1980), provides analytical expression for the terms $\sigma_{kl,ij}$:

A case of non-overlapping pair of asset returns ($(i, j) \neq (k, l)$ and $\{i, j\} \cap \{k, l\} = \emptyset$)

$$\begin{aligned} \sigma_{kl,ij} &= \frac{1}{2} \{ (\rho_{ki} - \rho_{kl} \rho_{li})(\rho_{lj} - \rho_{li} \rho_{ij}) + (\rho_{kj} - \rho_{ki} \rho_{ij})(\rho_{li} - \rho_{kl} \rho_{ki}) + \\ &+ (\rho_{ki} - \rho_{kj} \rho_{ij})(\rho_{lj} - \rho_{kl} \rho_{ij}) + (\rho_{kj} - \rho_{kl} \rho_{ij})(\rho_{li} - \rho_{lj} \rho_{ij}) \} \end{aligned}$$

A case of one common asset return in the two pairs (can be directly derived from the Equation above) (12)

$$\sigma_{kl,kj} = \rho_{lj}(1 - \rho_{kl}^2 - \rho_{kj}^2) - \frac{1}{2} \rho_{kl} \rho_{kj} (1 - \rho_{kl}^2 - \rho_{kj}^2 - \rho_{lj}^2)$$

A case of variance for (can be directly derived from the Equation above)²⁰

$$\sigma_{kl,kl} = (1 - \rho_{kl}^2)^2$$

Equations (11) and (12) allows us to comment on impact of portfolio size on the precision of average correlation estimate. First, we can compare precision of the average correlation estimate to the average precision of pairwise correlations:²¹

$$std_T(\widehat{\rho}_{Avg}) \leq \sum_{1 \leq k < l \leq K} w_{kl} std_T(\widehat{\rho}_{kl}) \quad (13)$$

Equation (13) shows that pooling the correlation estimates using average correlation in a portfolio context can reduce the uncertainty we typically see in individual pairwise correlations; certain portions of the statistical errors in the individual correlations are diversified away when considered within a portfolio.

¹⁹ The term under the arrow indicates speed of convergence.

²⁰ This formula represents asymptotic variance of $\widehat{\rho}_{kl}$, while Equation (1) provides asymptotic variance for the Fisher's z-transformation of $\widehat{\rho}_{kl}$

²¹ Equation (13) follows from basic properties of covariances: $\sigma_{kl,ij} \leq \sqrt{\sigma_{kl,kl}} \sqrt{\sigma_{ij,ij}}$. In that formula, we use notation $std_T(\widehat{\rho}_{kl}) = \sqrt{\frac{1}{T}} \sigma_{kl,kl}$.

How does the uncertainty change when we increase the size of the portfolio: $K \rightarrow \infty$? As Equation (12) indicates, it is possible that all cross-covariances of correlation estimates, $\sigma_{kl,ij}$, are not zero. This means that all the terms in Equation (11) could be non-zero (the total number of terms is $\left[\frac{K(K-1)}{2}\right]^2$). We assume the portfolio size increases in such a way that no sets of exposures dominate, meaning that $w_{kl} \rightarrow 0$ at rate $\frac{1}{K^2}$ as the portfolio size increases. In this case,

$$std_T(\widehat{\rho_{Avg}}) \rightarrow c \text{ as } K \rightarrow \infty, \text{ where } c \geq 0 \quad (14)$$

The asymptotic uncertainty c can be zero in some situations, such as when the true pairwise correlations ρ_{kl} are zero (or one). That is not the case in usual credit risk applications. The uncertainty in that context therefore cannot be arbitrarily reduced by increasing the size of the portfolio. In other words, there is a positive limit to which the statistical error can be mitigated in a large portfolio, but not beyond that limit.

The interpretation of the limit c is that it represents a systematic component of the statistical errors and it cannot be diversified away even in the large portfolio context. This systematic component exists as long as the true pairwise correlations are positive.

To understand the dynamics in Equation (14) better, we consider a simple case of a portfolio consisting of equally weighted exposures with equal characteristics and correlations: $w_{kl} = \frac{2}{K(K-1)}$ and $\rho_{kl} = \rho$ (for all $1 \leq k < l \leq K$):

$$\begin{aligned} std_T(\widehat{\rho_{Avg}}) &= \sqrt{\frac{1}{T}} \cdot \sqrt{\sum_{1 \leq k < l \leq K} \sum_{1 \leq i < j \leq K} w_{kl} w_{ij} \sigma_{kl,ij}} \\ &= \sqrt{\frac{1}{T}} \cdot w_K \cdot \sqrt{g_{1,K}(1-\rho^2)^2 + g_{2,K} \left[\rho(1-2\rho^2) - \frac{1}{2}\rho^2(1-3\rho^2) \right] + g_{3,K} [2\rho^2(1-\rho)^2]} \end{aligned}$$

where²²

$$w_K = \frac{2}{K(K-1)} \approx \frac{1}{K^2} \text{ is the weight of one instrument}$$

$$g_{1,K} = \frac{K(K-1)}{2} \approx K^2 \text{ is the number of variance terms}$$

$$g_{2,K} = \frac{2K(K-1)(K-2)}{2} \approx K^3 \text{ is the number of covariance terms with one overlapping asset return}$$

$$g_{3,K} = \frac{K(K-1)(K-2)(K-3)}{4} \approx K^4 \text{ is the number of covariance terms with no overlapping asset returns}$$

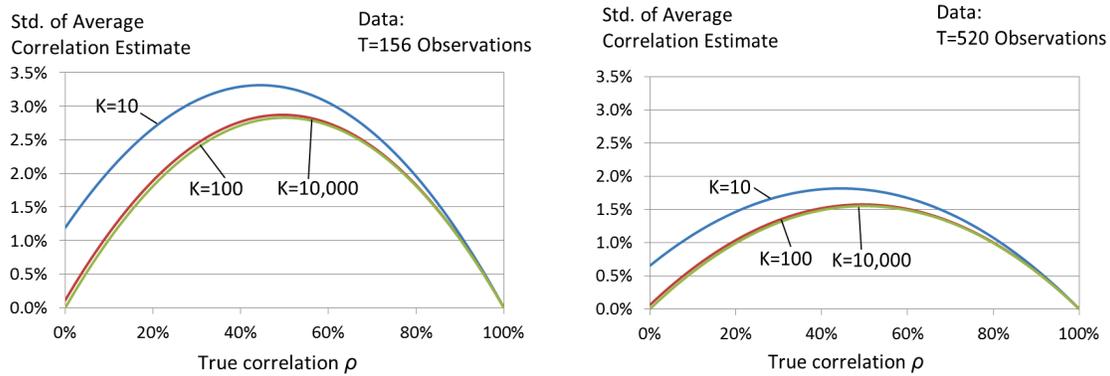
For a large portfolio, only the term $g_{3,K}$ matters, because only the number of number of covariance terms with no overlapping asset returns is large enough to outweigh the impact of diminishing weights w_K . Thus,²³

$$std_T(\widehat{\rho_{Avg}}) \rightarrow \sqrt{\frac{2}{T}} \cdot \rho(1-\rho) \text{ as } K \rightarrow \infty$$

²² The approximation \approx applies to large portfolio cases.

²³ We assume that $\rho \geq 0$.

Figure 3 Uncertainty in average correlation estimate under various parameters (number of observations T , portfolio size K , and level of correlation ρ), assuming that all pairwise correlations in the portfolio are equal.



Equation (15) provides an analytical expression for uncertainty in average correlation estimate, within a general case as well as in the case of a large portfolio. Figure 3 summarizes the uncertainty properties. As mentioned before, the number of observations T can potentially reduce the uncertainty by a substantial amount. On the other hand, Figure 3 shows that portfolio size has an impact when the portfolio size grows to tens of firms, but after it crosses one hundred firms, there is no further reduction in the uncertainty, because the correlation error components that are not systematic have already been diversified away.

For dependence on correlation levels, as Equation (16) and Figure 3 suggest, the estimates exhibit minimum error for true correlation levels around zero and one, while the uncertainty is largest for correlation levels around 50%.

We have focused so far on the simple case with equal correlations. We next show an example of realistic portfolios containing 10 and 500 large corporate exposures across multiple countries and industries. Table 1 summarizes distributions of pairwise correlations in the portfolios.

TABLE 1

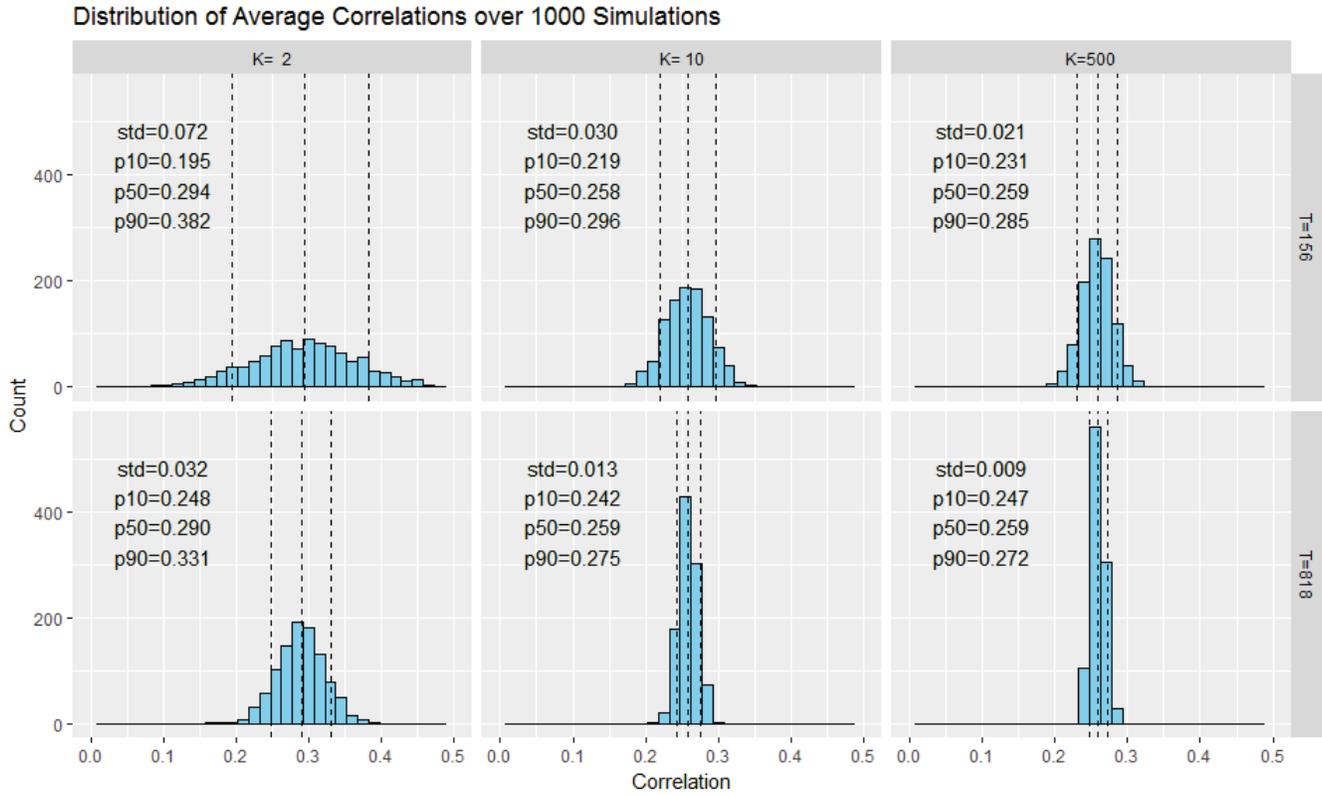
Correlation Summary for Sample Portfolios

PORTFOLIO	ASSET CORRELATION SUMMARY STATISTICS						
	MEAN	STANDARD DEVIATION	5 TH PERCENTILE	25 TH PERCENTILE	MEDIAN	75 TH PERCENTILE	95 TH PERCENTILE
10 Exposures	26.0%	4.6%	19.9%	22.4%	25.5%	28.8%	35.3%
500 Exposures	26.0%	9.5%	12.3%	19.3%	24.9%	31.5%	43.0%

Figure 4 presents a simulation analysis, where we simulate time series of asset returns from the correlations presented in Table 1 over $T = 156$ weeks and 818 weeks.²⁴ Subsequently, we summarize the sampling distribution of the average correlation across the simulation trials. As Figure 4 shows, increasing the number of observations substantially reduces the variance of the distribution. While the variance of average correlation for the portfolio of 10 exposures is considerably lower than if we assume the average correlation is based on a single pair of returns, the magnitude of improvement falls when we move to the 500 exposure portfolio. Larger portfolios would exhibit even less marginal improvement, as discussed earlier in this section.

²⁴ The 818 observations corresponds to the number of observations in GCorr 2015 Corporate, where we use weekly data from July 1, 1999–March 31, 2015.

Figure 4 Simulated distribution of average correlation estimates under various parameters (number of observations T , portfolio size K).



3.2 Factor Model

Section 3.1 examines the case of estimating correlations as pairwise sample correlations. In practice, however, correlations are often estimated using a factor model. Let us therefore consider a factor model

$$r_k = \sqrt{RSQ_k} \cdot \phi_k + \sqrt{1 - RSQ_k} \cdot \varepsilon_k, \quad (16)$$

where ϕ_k is the systematic factor for firm k , a linear combination of an underlying set of several factors (for example, representing country and industry risk), and ε_k is the idiosyncratic factor for firm k . Parameter RSQ_k reflects the sensitivity of the firm k 's asset return to its systematic factor – $corr(r_k, \phi_k) = \sqrt{RSQ_k}$. Asset correlations across firms in the factor model can be expressed as:

$$\rho_{kl} = corr(r_k, r_l) = \sqrt{RSQ_k} \sqrt{RSQ_l} \cdot corr(\phi_k, \phi_l) \quad (17)$$

We focus specifically on a single-factor model, defined as $\phi_k = \phi$, which allows us to study uncertainty in correlation estimates using analytical methods in the single-factor model, $\rho_{kl} = \sqrt{RSQ_k} \sqrt{RSQ_l}$

Estimating correlations across firms thus means estimating the RSQ values,²⁵ accomplished by regressing time series of asset returns on the time series of the factor and reporting \widehat{RSQ}_k — the regression RSQ values (coefficients of determination).

Given that $\sqrt{RSQ_k}$ is equivalent to the empirical correlation between the asset return and systematic factor time series, we can use results from Section 3.1 to describe statistical properties of correlation estimates in the single-factor model. We first focus on

²⁵ We assume that the time series of the factor ϕ is available. In practice, this factor can be unobservable, in which case, we must estimate it from the time series of the asset returns. This process can add error to the estimates of correlations across firms. Such an error will be small, however, if the factor is estimated based on asset returns of a large number of firms.

the estimator of a single correlation between two firms: $\widehat{\rho}_{kl} = \sqrt{RSQ_k} \cdot \sqrt{RSQ_l}$. We can use the delta method to quantify the uncertainty in this estimator measured by the standard deviation:²⁶

$$\begin{aligned} std_T(\widehat{\rho}_{kl}) &= \\ &= \sqrt{RSQ_l \cdot var_T(\sqrt{RSQ_k}) + RSQ_k \cdot var_T(\sqrt{RSQ_l}) + 2\sqrt{RSQ_k}\sqrt{RSQ_l} \cdot cov_T(\sqrt{RSQ_k}, \sqrt{RSQ_l})} \\ &= \sqrt{\frac{1}{T} \cdot \sqrt{(1 - RSQ_k RSQ_l)^2 - (1 - RSQ_k)(1 - RSQ_l)}} \end{aligned} \quad (18)$$

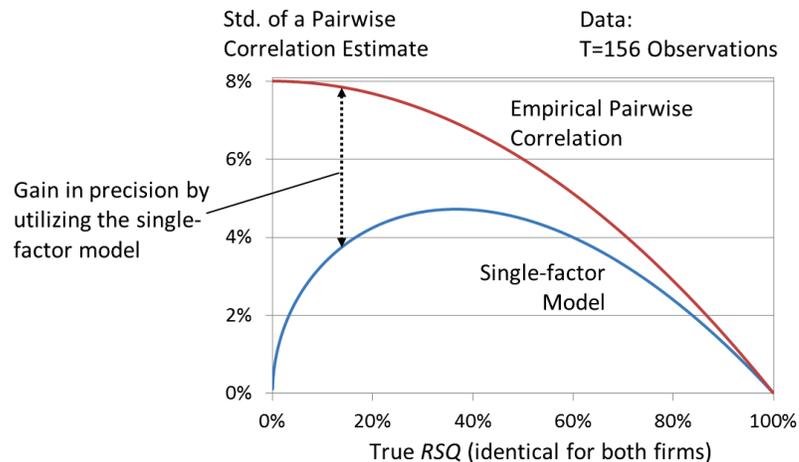
We provide a detailed derivation of Equation (18) in Appendix A. What if the correlation $\widehat{\rho}_{kl}$ is not estimated using the single-factor model, but simply as an empirical pairwise correlation of the asset return time series? In that case, the standard deviation of the estimate can be obtained from Equation (12), while using the link between the correlation and RSQ values in Equation (17):

$$std_T^{pairwise}(\widehat{\rho}_{kl}) = \sqrt{\frac{1}{T} \cdot \sqrt{(1 - RSQ_k RSQ_l)^2}} \quad (19)$$

We compare the uncertainty in an individual correlation estimate based on the single-factor model, $std_T(\widehat{\rho}_{kl})$ from Equation (18), and based on the pairwise empirical correlation, $std_T^{pairwise}(\widehat{\rho}_{kl})$ from Equation (19). It is clear that the single-factor model estimate contains less statistical error, which leads us to the following conclusion: if asset returns follow a single-factor model, it is preferable to utilize that structure in estimation as opposed to calculating empirical pairwise correlations.

We illustrate this precision gain using the single-factor model in Figure 5, where we plot the standard deviations for the two cases in relation to RSQ_k (we set $RSQ_l = RSQ_k$) in that example. As the chart shows, the single-factor model enables the largest benefits in reducing statistical errors in cases of low RSQ values. Even for RSQ values of around 50%, however, the statistical error is reduced by a quarter. Given that the range from low to 50% RSQ values covers typical corporates (for example, based on the GCorr Corporate Model), these precision gains help with estimation of correlations in practice. The benefit of the single-factor model vanishes with RSQ values approaching 100%.

Figure 5 Uncertainty in a pairwise correlation estimate under a single-factor model assumption with the two firms having identical RSQ values.



We have focused on one pairwise correlation in a single-factor model so far. Next, we explore statistical properties of the average correlation estimate for a portfolio, $\widehat{\rho}_{Avg} = \sum_{1 \leq k < l \leq K} w_{kl} \cdot \widehat{\rho}_{kl}$ where $\widehat{\rho}_{kl}$ is determined within the single-factor framework. To obtain tractable analytical expressions, we consider the asymptotic case of a large, equally-weighted portfolio, with equal correlations ($K \rightarrow \infty$, $w_{kl} = w_K$, and $RSQ_k = RSQ$). As Equation (15) indicates, the uncertainty in $\widehat{\rho}_{Avg}$ is driven only by

²⁶ Due to the assumptions of the delta method, the formula applies only if $RSQ_k > 0$ and $RSQ_l > 0$.

covariances of correlation estimates of distinct pairs of firms ($\sigma_{kl,ij}$ for $(i, j) \neq (k, l)$ and $\{i, j\} \cap \{k, l\} = \emptyset$). In Appendix A, we derive expressions for these covariances within the single-factor framework, and show that:

$$\text{std}_T(\widehat{\rho}_{Avg}) \rightarrow \sqrt{\frac{2}{T}} \cdot RSQ \cdot (1 - RSQ) \text{ as } K \rightarrow \infty \quad (20)$$

The question is how this uncertainty compares to the case when the correlations $\widehat{\rho}_{kl}$ are estimated as pairwise empirical correlations. By using the result from Equation (15) and recognizing that $\rho = RSQ$ (pairwise correlation equals the RSQ value in a single-factor model with identical RSQ values across exposures), we obtain that $\text{std}_T^{\text{pairwise}}(\widehat{\rho}_{Avg})$ converges to $\sqrt{2/T} \cdot RSQ \cdot (1 - RSQ)$ — an identical asymptotic value to the one in Equation (20). This means that using a single-factor model does not improve portfolio-wide statistical performance for large granular portfolios. In other words, the portfolio-wide statistical error hits the same lower boundary for both the empirical pairwise correlation approach and the single-factor approach, and there cannot be additional precision gains.²⁷

What are the implications of the results from this section? Compared to using empirical pairwise correlations as an estimation method, a factor model (a single-factor model in our examples) reduces statistical error in estimates of individual pairwise correlations but not in the portfolio-level, average correlation statistic, if the portfolio is large and well-diversified. Under the circumstances considered in our examples, there is therefore no improvement in statistical quality of the overall credit portfolio risk measure using a factor model for such a portfolio. The improvement for individual pairwise correlations is, however, very important, because a credit portfolio analysis does not involve only calculating portfolio-level measures for large well-diversified portfolio. Some portfolios are heavily concentrated in several names, or contain only few exposures — and that is where precision in individual pairwise correlations matters.

Using a factor model provides additional benefits. This paper assumes an idealized case utilizing asset return time series of the same length and quality for all firms. This is rarely the case in practice. Factor models are better equipped to deal with data issues than calculate empirical pairwise correlations, which makes using appropriately defined factor models preferable to relying on (large) matrixes of empirical pairwise correlations.

In order to illustrate the results of this section on a realistic portfolio, we take the portfolios from Table 1, and simulate asset returns to calculate a sampling distribution of average correlations, as well as a select pairwise correlation. We perform this exercise, however, under three settings: estimating the correlations as pairwise correlations, as correlations coming from a single-factor model, and as correlations coming from a multi-factor model.²⁸ To ensure that the correlations can be represented as coming from a single- or a multi-factor model, we adjust the values summarized in Table 1 as necessary, while keeping the mean the same as the original mean value in the table.

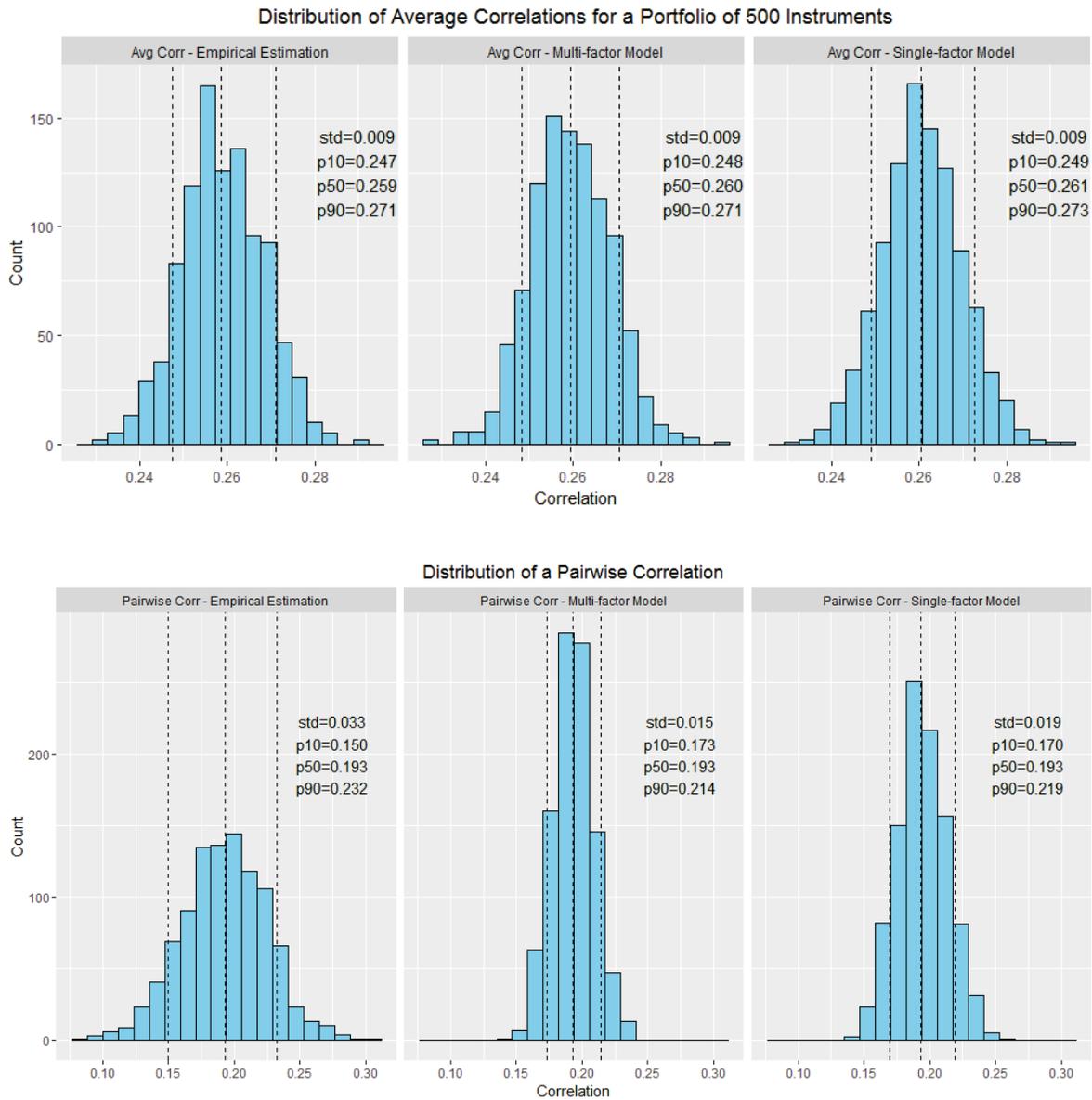
As Figure 6 shows, the average correlation distributions based on the three methods are similar. On the other hand, the correlation estimate for a single pair of asset returns shows less uncertainty if based on a factor model rather than calculated as an empirical pairwise correlation.

Note, in the multi-factor model specified in (17), the correlation estimate consists of estimated RSQ value as well as estimated factor correlation $\text{corr}(\widehat{\phi}_k, \widehat{\phi}_l)$. Given the presence of estimated factor correlations and given the differences in RSQ, uncertainty in the multi-factor model case and the single-factor model case should not be compared.

²⁷ As in Equation (15), the other terms besides $\sigma_{kl,ij}$ for distinct pairs matter if the portfolio is small or not granular. In that case, the single-factor model can bring some reduction in the portfolio-level statistical error.

²⁸ We simulate 110 factors (49 country factors and 61 industry factors) and assume that all factors ϕ_k in Equation (16) are linear combinations of these 110 factors.

Figure 6 Distribution of average and pairwise correlation estimates under: empirical correlation estimation, a single-factor model, a multi-factor model assumption. We use the portfolio of 500 exposures from Table 1.



4. Impact of Uncertainty in Correlations on Risk Measure Estimates

This section presents numerical examples illustrating the magnitude of uncertainty in portfolio risk measures. We also examine the results within context of the previous discussion because the uncertainty at the portfolio level can be conceptually inferred from uncertainty in the average correlations, which we demonstrated in Section 2. We then addressed the uncertainty in the average correlations in Section 3.

We consider a credit model where loss is incurred only in the case of a default. As in Section 2.1, a default occurs only if the borrower's asset return falls below a pre-specified barrier, given by the borrower's credit quality: $r_k < N^{-1}(PD_k)$. If default occurs, the loss is a fixed proportion of notional, LGD_k . The portfolio loss is therefore given by the following formula:

$$\tilde{L} = \sum_{k=1}^K w_k \cdot LGD_k \cdot I_{[r_k < N^{-1}(PD_k)]} \quad (21)$$

The random changes in credit qualities r_k drive the randomness in loss L . Note, we can express the expected loss as $EL = \sum_{k=1}^K w_k \cdot LGD_k \cdot PD_k$.

4.1 Portfolio-Level Risk Measures

Equation (21) serves as a starting point for calculating the two portfolio level risk measures: Unexpected Loss as $UL = std(\tilde{L})$ and Economic Capital as the percentile of the loss in excess of expected loss: $P(\tilde{L} - EL > EC) = \alpha$.

We begin by focusing on the UL , which we can express analytically as follows:

$$UL = \sqrt{\sum_{k=1}^K w_k^2 UL_k^2 + \sum_{\substack{k,l=1 \\ k \neq l}}^K w_k w_l LGD_k LGD_l \cdot \{N_2(N^{-1}(PD_k), N^{-1}(PD_l), \rho_{kl}) - PD_k PD_l)\}} \quad (22)$$

where $UL_k = LGD_k \sqrt{PD_k(1 - PD_k)}$, symbol N_2 denotes the bivariate standard normal cumulative distribution function, and ρ_{kl} is the correlation of asset returns r_k and r_l .

Equation (22) applies to all correlation models, which enter the formula through the asset correlation ρ_{kl} . In order to assess impact of uncertainty in correlations on unexpected loss, we study the distribution of \widehat{UL} — the value of UL from Equation (22), which uses estimated parameters, $\widehat{\rho}_{kl}$, in place of asset correlations.

In addition to unexpected loss, we also analyze the impact of uncertainty in correlations on economic capital. Given that there is no analytical formula for EC , bar very special cases, we estimate this risk measure using simulation. We run this simulation with estimated correlation parameter, $\widehat{\rho}_{kl}$, and obtain \widehat{EC} simply as the quantile of the simulated distribution of \tilde{L} . Thus, the uncertainty in the economic capital captures both uncertainty in correlation estimates as well as sampling error due to simulation. The only case where capital can be expressed analytically is the single-factor case:

$$EC = \sum_{k=1}^K w_k \cdot LGD_k \cdot \left\{ N \left(\frac{N^{-1}(PD_k) - \sqrt{RSQ_k} \cdot N^{-1}(\alpha)}{\sqrt{1 - RSQ_k}} \right) - PD_k \right\} \quad (23)^{29}$$

We denote by \widehat{EC} , the single-factor capital calculated using Equation (23), but with estimated RSQ values \widehat{RSQ}_k . Note, the randomness in this capital measure is driven completely by statistical error in the RSQ value estimate.

Equation (9) in Section 2 linked credit risk measures to the properties of average correlations, which we discussed in Section 3. We can now apply the conclusions presented therein to estimates \widehat{UL} and \widehat{EC} .

We consider two portfolios, equally weighted with 10 and 500 exposures, asset correlations from Table 1, PD s defined as EDF s of the 10 and 500 largest global corporates, and LGD s set to 40%. Figure 7 and Figure 8 present the magnitude of uncertainty for the portfolios and for the various risk statistics. In our simulation analysis, even for a portfolio with 500 exposures, \widehat{EC} is a

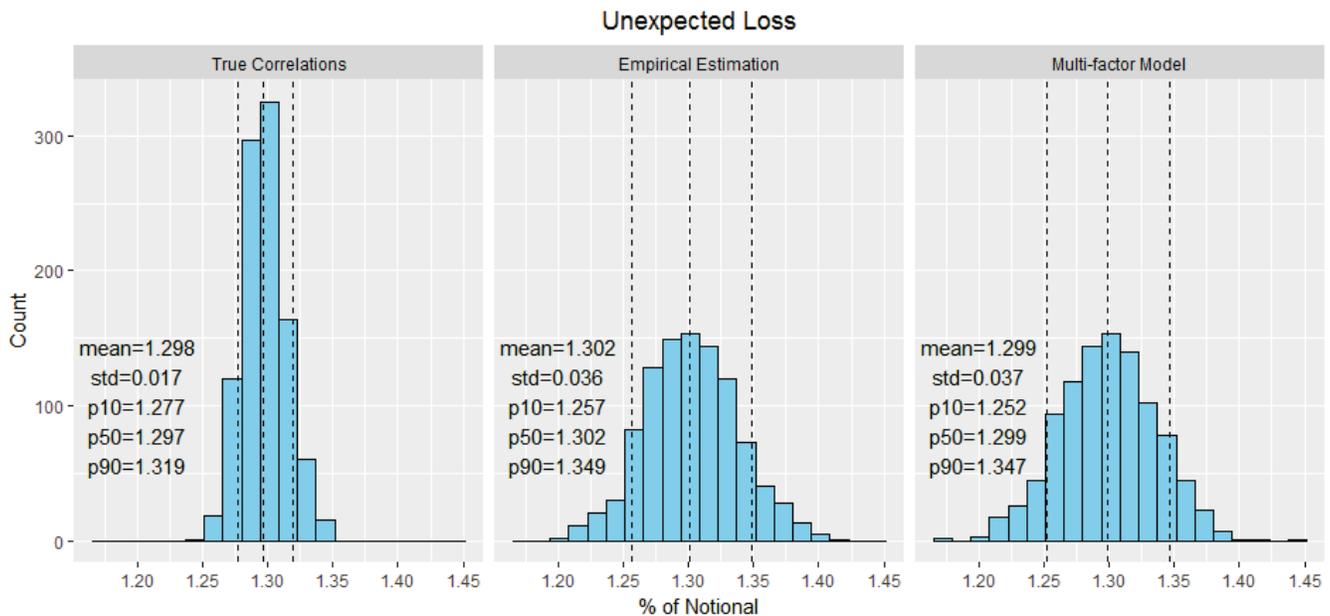
²⁹ We can use the formula for EC in this case to illustrate the term $\xi = \left| \sum_{1 \leq k < l \leq K} \frac{\partial f}{\partial \rho_{kl}} \right|$ from Equation (9). Taking $RSQ_k = \rho_{k\phi}$ as the correlation of an obligor's asset return with the systematic factor and interpreting the Economic Capital in (23) as $EC = \sum_{1 \leq k \leq K} \frac{1}{K} \cdot h(RSQ_k)$, the term ξ becomes $\left| \sum_{1 \leq k \leq K} \frac{1}{K} \cdot \frac{\partial h}{\partial RSQ_k} \right|$. Assuming equal weights for exposures in the portfolio, and equal PD, LGD, and RSQ values, we can write the term as $\left| \sum_{1 \leq k < l \leq K} \frac{1}{K} \cdot \frac{\partial h}{\partial RSQ} \right| = \frac{\partial h}{\partial RSQ}$. Thus, the uncertainty in the Economic Capital due to correlation error is $std_{\tau}(\widehat{EC}) = \frac{\partial h}{\partial RSQ} \cdot std_{\tau}(\widehat{\rho}_{Avg})$.

discretely distributed variable concentrated in only several values (for example, four values) — these patterns are visible in Figures 7 and 8.

As we can see, the differences between uncertainties in the portfolio-level risk measures calculated based on the empirical pairwise correlations and factor models are not substantial.³⁰

Figure 7 also illustrates the magnitude of error in the estimated portfolio-level risk measures. As an example, the range between the 10th and 90th percentile for 10bps capital is approximately 24bps (2.80%–3.04%), and the standard deviation of the capital is 12bps.³¹ It is worth highlighting that this range and the standard deviation include two effects: the uncertainty in correlation estimates and the simulation noise from estimating the capital. We therefore plot a histograms in Figure 7, where we isolate the simulation noise impact by assuming constant (“true”) correlations across all trials. As the results indicate, of the 24bps 10th-90th percentile range, 16bps come from the simulation noise, and the remaining part from the correlation estimate uncertainty. Similarly, 7bps of the capital volatility of 12bps comes from the simulation noise, while the remaining part can be attributed to the correlation estimate uncertainty.

Figure 7 Uncertainty in \widehat{UL} , \widehat{EC} for the portfolio with 500 exposures. We consider the empirical pairwise correlation method and the multi-factor model for estimating correlations. We use a simulation method to estimate \widehat{EC} and, therefore, present its distribution if we use the true correlations to illustrate how much simulation noise contributes to its uncertainty.



³⁰ This can be shown using statistical tests. With the knowledge that not all assumptions of the F-tests for comparing volatilities or the Kolmogorov-Smirnov test for comparing distributions are met, we still perform the tests to obtain a general indication of whether the differences in volatilities and distributions are not substantial. In line with the theoretical results presented in this paper, the tests do not detect significant differences between uncertainties in the portfolio risk measures based on the empirical and factor-based estimations.

³¹ Note, the portfolio's PD level impacts the capital level: the average PD for the portfolio with 500 exposures is 33.5 bps.

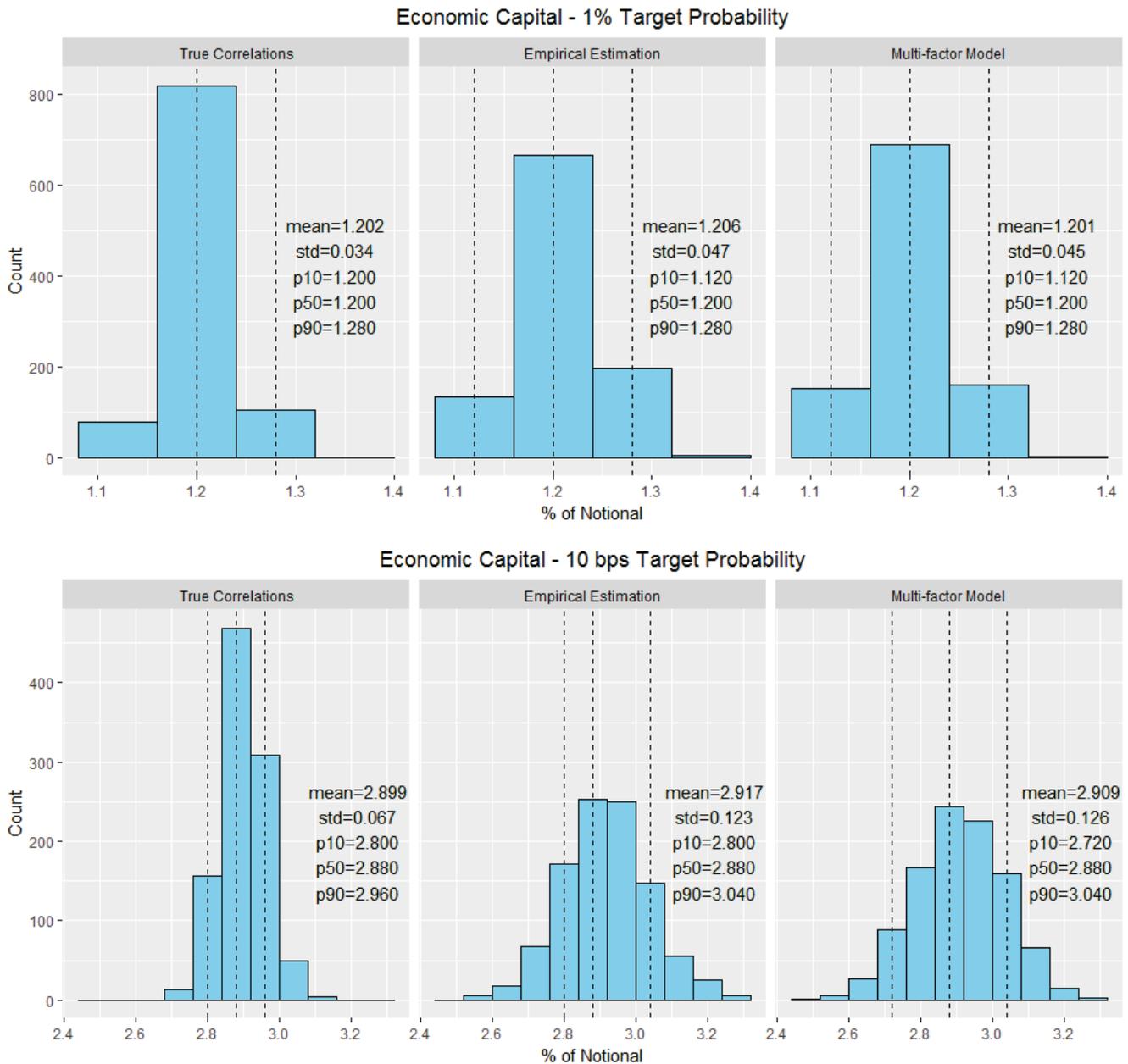
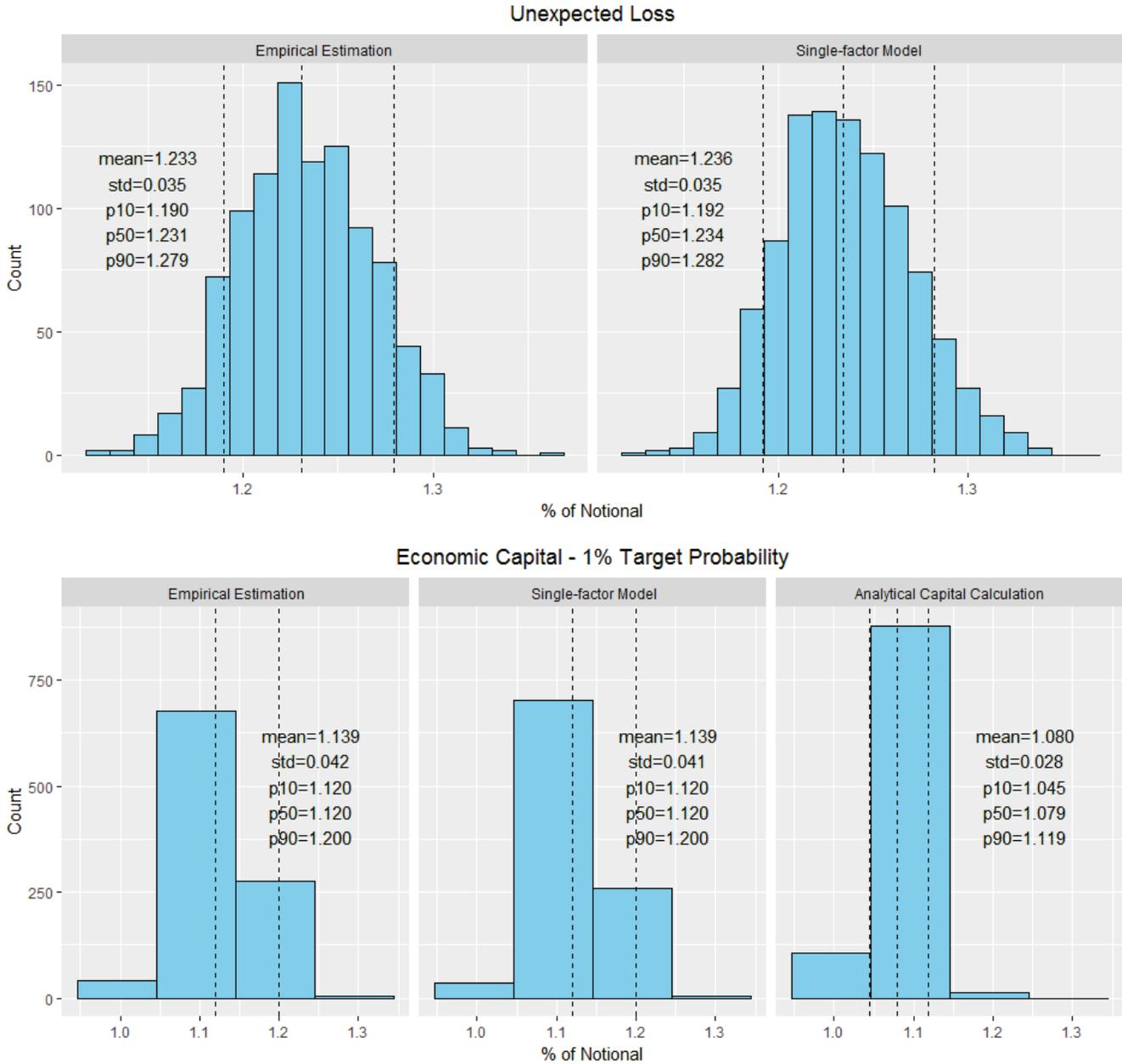


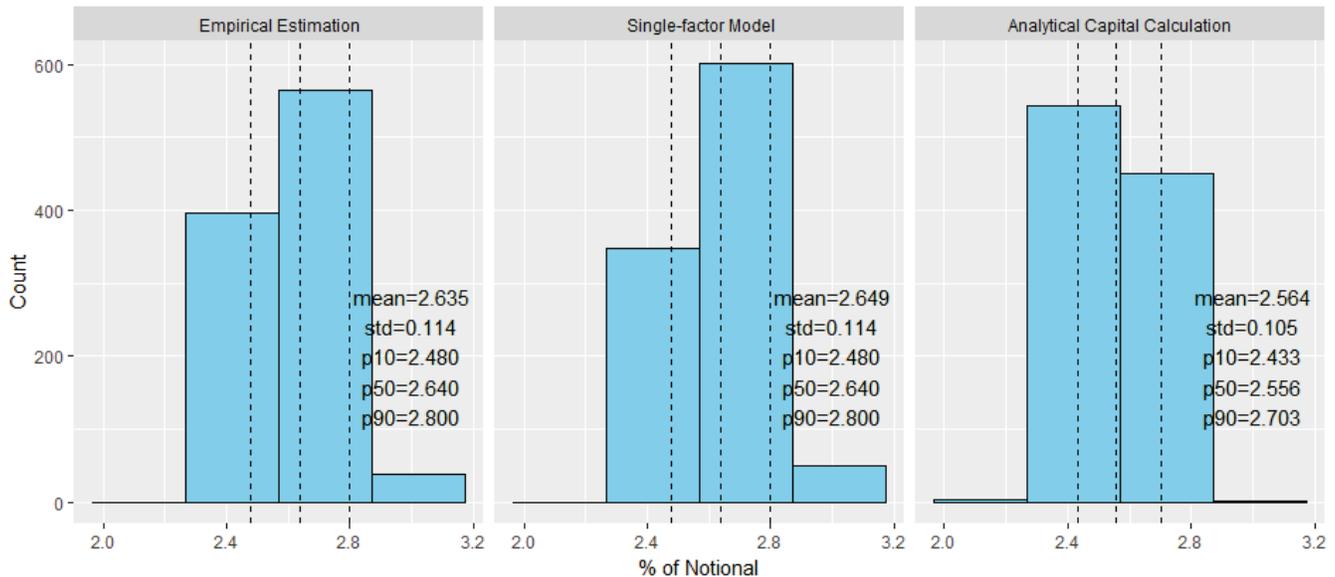
Figure 8 contrasts risk statistics based on the empirical correlation and single-factor approaches for estimating correlations. Conclusions are similar to those in Figure 7 — the uncertainty of portfolio-level statistics do not differ substantially. Thanks to Equation (24), we can determine \bar{EC} analytically for the single-factor case and thus remove the simulation noise effect. We show results based on this analytical capital calculation in the third histograms of Figure 8.³²As the histogram with the analytical calculation shows, the range between the 10th percentile and 90th percentile of the 10bps \bar{EC} in the single-factor case is 2.43%-2.70%, thus 27bps.

³² Note, the mean capital is different for the simulation case and analytical calculation case. The reason is the aforementioned discretization effect in the simulation case.

Figure 8 Uncertainty in \widehat{UL} , \widehat{EC} for the portfolio with 500 exposures. We consider the empirical pairwise correlation method and the single-factor model to estimate correlations. While \widehat{EC} for the empirical pairwise correlation and for the chart labeled as single-factor model is subject to noise (it is estimated using simulation), the uncertainty in \widehat{EC} in the analytical calculation case is purely due to correlation uncertainty within the single-factor model.



Economic Capital - 10 bps Target Probability



4.2 Decomposition of Uncertainty in a Multi-Factor Model

As Equation (17) indicates, the correlation estimate in a multi-factor model is calculated as $\widehat{\rho}_{k,l} = \sqrt{RSQ_k} \sqrt{RSQ_l} \cdot \text{corr}(\widehat{\phi}_k, \widehat{\phi}_l)$. We applied this formula in the multi-factor cases in Section 4.1. This section presents an example that decomposes the total impact of uncertainty in $\widehat{\rho}_{k,l}$ on portfolio risk measure into two parts — one attributed to the effects of uncertainty in RSQ values and one to the uncertainty in the factor correlations $\text{corr}(\widehat{\phi}_k, \widehat{\phi}_l)$.

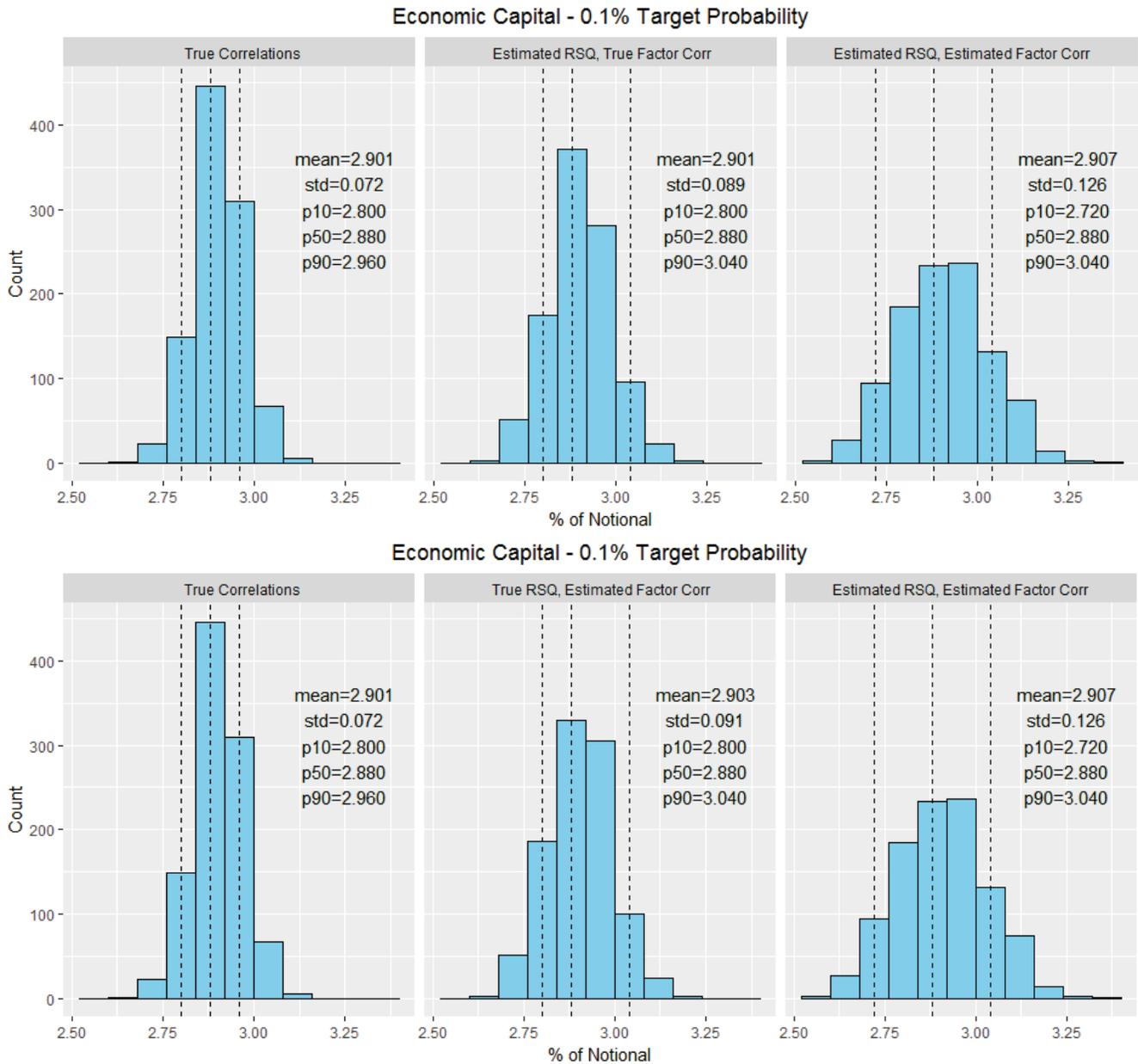
We consider the portfolio of 500 exposures from Table 1 and run three simulations, similar to those in Figure 7, to obtain the distribution of \widehat{EC} (with 10bps tail probability). The first simulation assumes no uncertainty in any correlation parameter (and the distribution thus reflects simulation noise in the capital estimate). The second simulation adds the effect of uncertainty in RSQ; and the third simulation considers the effects of joint uncertainty in RSQ and factor correlations, equivalent to considering the overall uncertainty in asset correlations using a multi-factor model. Acknowledging that results of the exercise depend on sequencing of the runs, we conduct another exercise in which we add uncertainty in factor correlations in the second run and then add uncertainty in RSQ in the third run.

We are interested in the marginal increases in uncertainty in the capital estimate between the first and second runs, and then between the second and third runs. Figure 9 summarizes the results of the exercise.

As results suggest, the RSQ and factor correlation uncertainty effects are similar. One reason is that the average factor correlation in this exercise is 58%, not substantially different from the average RSQ value of 46% — thus, the two components in $\sqrt{RSQ_k} \sqrt{RSQ_l} \cdot \text{corr}(\widehat{\phi}_k, \widehat{\phi}_l)$ are on average roughly similar. If factor correlations were substantially higher³³, for example, the effect of RSQ uncertainty would dominate the factor correlations uncertainty.

³³ The average factor correlation of 58% in our portfolio reflects the fact that exposures are dispersed across multiple countries. If all exposures belonged to a single country, the average factor correlation — as implied by the GCorr model — would be much higher, closer to the level of 80%. In that case, the factor correlations would capture only diversifications across industries, but not across countries.

Figure 9 \widehat{EC} distribution for the portfolio with 500 exposures under the multi-factor correlation assumption. On the first row: the first histogram assumes no uncertainty in the correlation parameters; the second histogram adds the effect of uncertainty in RSQ; the third histogram considers the effects of joint uncertainty in RSQ and factor correlations. The second row of histograms adds the effect of uncertainty in factor correlations first, and uncertainty in RSQ afterwards.



4.3 Portfolio-Referent Exposure-Level Risk Measures

Sections 4.1 and 4.2 focus on portfolio-level risk measures. The question we examine here is how the uncertainty in correlations impacts exposure-level (but portfolio-referent) risk measures. Such measures include Risk Contribution (RC) and Tail Risk Contribution (TRC), widely used for allocating the overall portfolio risk to individual exposures.³⁴ For an exposure k , the measures

³⁴ More details on the concepts of Risk Contribution and Tail Risk Contribution can be found in Levy (2008): *An Overview of Modeling Credit Portfolios*, and in the Moody's Analytics document: *Modeling Credit Portfolios*.

are formally defined as sensitivities of the portfolio's Unexpected Loss and Economic Capital to the exposure's weight in the portfolio: $RC_k = \frac{\partial UL}{\partial w_k}$ and $TRC_k = \frac{\partial EC}{\partial w_k}$.

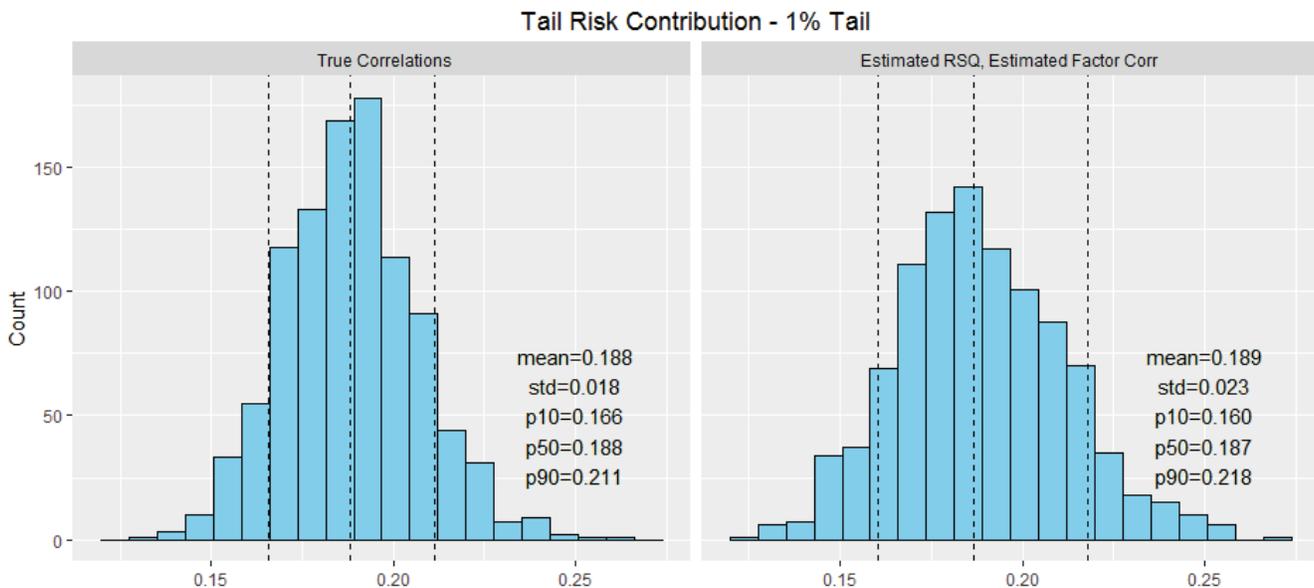
Appendix B shows that RC_k represents a (scaled) correlation between the value of the exposure and the portfolio's remaining value, which, in turn, depends on individual asset correlations within the entire portfolio. Thus, uncertainty in those correlations translates into a Risk Contribution estimate with a statistical error: \widehat{RC}_k . In Appendix B, we also demonstrate that the statistical error in \widehat{RC}_k within a large portfolio depends on uncertainty in the overall portfolio average correlation ($\widehat{\rho}_{Avg}$), as well as on uncertainty in the average asset correlation of exposure k with the rest of the portfolio: $\widehat{\rho}_{Avg,k} = \frac{1}{K} \sum_{l=1}^K w_l \cdot \widehat{\rho}_{k,l}$.

Under simplified circumstances (equal correlations), and for a large portfolio, $\widehat{\rho}_{Avg,k}$ has similar properties as $\widehat{\rho}_{Avg}$, as we discuss in more detail in Appendix B — namely that its uncertainty remains the same, whether estimating correlations using a factor model or as empirical correlations. The reasoning is similar to the case of average correlation estimate $\widehat{\rho}_{Avg}$: the individual correlation errors that the factor model performs well at controlling are eliminated by averaging. Thus, the factor model loses its advantage over the purely empirical factor estimate in a large well-diversified portfolio. As a result, the magnitude of uncertainty in \widehat{RC}_k remains the same, irrespective of whether we use empirical correlation estimates or a factor model.

We reach similar conclusions using simulation analyses, if we assume realistic portfolio (i.e. correlations not equal), and if we consider TRC .

To illustrate the magnitude of the uncertainty in TRC , we present an example of an instrument in Figure 10. The two histograms compare the \widehat{TRC}_k uncertainty due to simulation noise only and due to simulation noise with the uncertainty in correlations. As the results show, the 10th-90th percentile range in \widehat{TRC}_k shows an increase from 17-21bps to 16-22bps as we introduce uncertainty in correlations. The standard deviation increases from 1.8bps to 2.3 bps.

Figure 10 \widehat{TRC}_k distribution for the portfolio with 500 exposures, 10bps target probability, under the multi-factor correlation assumption. Instrument k has a PD of 1.02% and an RSQ of 50%.



Even though we conclude that uncertainty in the portfolio-level and exposure-level risk measures for large diversified portfolios does not depend on whether correlations are estimated as sample correlations or with a factor model, we should highlight the overall benefits of using factor models. Not only do they impose economic structure on correlations, thus allowing easier interpretation of the matrix or expert judgment overlay, they also control for outliers among individual pairwise correlations. Thus, they reduce statistical error when analyzing individual pairs of firms or when portfolios are either small or heavily concentrated in several names or categories.

5. Summary

This document describes how statistical errors in correlation estimates impact risk analysis of a credit portfolio. The statistical errors it focuses on represent sampling errors that naturally arise due to limited length of that time series which are used for estimation credit correlations among obligors. In line with portfolio context of our analysis, we do not examine single, pairwise correlation estimates in isolation (although we revisit well-known properties as a benchmark), but rather all pairwise correlations within the portfolio. The highly non-linear nature of the relationship between portfolio risk and correlations presents additional challenges for analyzing credit portfolios (compared to, for example, market portfolios), which we overcome by using the delta method.

We show that under certain assumptions, we can derive analytical properties of uncertainty in the portfolio risk measures. We link that uncertainty to uncertainty in a measure of average asset correlation. Our first conclusion is expected—uncertainty in the portfolio risk measures decreases with lengthening of the time series (at the rate of $1/\sqrt{T}$). Our second conclusion relates to portfolio size—as the number of obligors increases to a large number, reduction in the estimation error of average correlation, and, in turn, reduction in the estimation error of the portfolio risk measure, slows down, and the magnitude of the error remains positive, even for very large portfolios.

We examine effects of statistical errors coming from three methods of estimating correlations: empirical correlations, a single-factor model, and a multi-factor model. While we show the statistical errors in portfolio risk measures for very large portfolios are identical across these three methods, the factor models have several superior properties over empirical correlations. They impose economic structure on correlations and reduce the impact of outliers, as demonstrated by the lower uncertainty in individual pairwise correlation estimates under factor models. This effect is important for smaller portfolios or portfolios with heavy concentrations in several names.

We complement the conclusions with numerical examples that use portfolios with realistic correlation and PD parameters. These examples also illustrate the magnitude of the statistical errors in estimates of risk measures for those portfolios.

Appendix A Uncertainty in a Single-Factor Model

This appendix revisits several analytical results for a single-factor model to provide more details regarding derivation.

UNCERTAINTY IN A PAIRWISE CORRELATION

Equation (18) presents the standard deviation of the pairwise correlation in a single-factor model: $std_T(\widehat{\rho}_{kl})$. In such a model, the correlation estimate is a function of the RSQ values: $\widehat{\rho}_{kl} = \sqrt{RSQ_k \cdot RSQ_l}$. We can rewrite the RSQ value as $\sqrt{RSQ_k} = \widehat{\rho}_{k\phi}$ — an estimate of the correlation between obligor k 's asset return and the systematic factor ϕ . We denote the covariance matrix of $\begin{bmatrix} \widehat{\rho}_{k\phi} \\ \widehat{\rho}_{l\phi} \end{bmatrix}$ as $\Sigma_{k,l,\phi}$. Note, the matrix entries represent variances and the covariance of $\widehat{\rho}_{k\phi}$ and $\widehat{\rho}_{l\phi}$ as provided by Equation (12)³⁵.

We can then express the pairwise correlation estimate as $\widehat{\rho}_{kl} = g(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi})$, where $g(x, y) = x \cdot y$. If the delta method assumptions are met, $std_T(\widehat{\rho}_{kl})$ can be asymptotically approximated as³⁶ $\sqrt{\nabla g(\rho_{k\phi}, \rho_{l\phi})^T \times \Sigma_{k,l,\phi} \times \nabla g(\rho_{k\phi}, \rho_{l\phi})} = \sqrt{\rho_{l\phi}^2 \cdot var_T(\widehat{\rho}_{k\phi}) + \rho_{k\phi}^2 \cdot var_T(\widehat{\rho}_{l\phi}) + 2\rho_{k\phi}\rho_{l\phi} \cdot cov_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi})}$.

Applying Equation (12), and recognizing that $\rho_{k\phi} = \sqrt{RSQ_k}$, we obtain the result from Equation (18), for a sufficiently large T :

$$\begin{aligned} std_T(\widehat{\rho}_{kl}) &= \sqrt{\rho_{l\phi}^2 \cdot var_T(\widehat{\rho}_{k\phi}) + \rho_{k\phi}^2 \cdot var_T(\widehat{\rho}_{l\phi}) + 2\rho_{k\phi}\rho_{l\phi} \cdot cov_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi})} \\ &= \sqrt{\frac{1}{T} \cdot \sqrt{\rho_{l\phi}^2(1 - \rho_{k\phi}^2)^2 + \rho_{k\phi}^2(1 - \rho_{l\phi}^2)^2 + 2\rho_{k\phi}^2\rho_{l\phi}^2 \left[(1 - \rho_{k\phi}^2 - \rho_{l\phi}^2) - \frac{1}{2}(1 - \rho_{k\phi}^2 - \rho_{l\phi}^2 - \rho_{k\phi}^2\rho_{l\phi}^2) \right]}} \\ &= \sqrt{\frac{1}{T} \cdot \sqrt{\rho_{l\phi}^2(1 - \rho_{k\phi}^2)^2 + \rho_{k\phi}^2(1 - \rho_{l\phi}^2)^2 + \rho_{k\phi}^2\rho_{l\phi}^2(1 - \rho_{k\phi}^2)(1 - \rho_{l\phi}^2)}} \quad (24) \\ &= \sqrt{\frac{1}{T} \cdot \sqrt{RSQ_l \cdot (1 - RSQ_k)^2 + RSQ_k \cdot (1 - RSQ_l)^2 + RSQ_k RSQ_l \cdot (1 - RSQ_k)(1 - RSQ_l)}} \\ &= \sqrt{\frac{1}{T} \cdot \sqrt{(1 - RSQ_k RSQ_l)^2 - (1 - RSQ_k)(1 - RSQ_l)}} \end{aligned}$$

COVARIANCE OF TWO PAIRWISE CORRELATION ESTIMATES

Equation (20) provides the standard deviation of the average correlation in a single factor model, $std(\widehat{\rho}_{Avg})$. As we show in Section 3.1, for a large portfolio, the only terms that matter for this standard deviation are covariances $\sigma_{kl,ij} = \sqrt{T} \cdot cov_T(\widehat{\rho}_{kl}, \widehat{\rho}_{ij})$. In a single-factor model, the pairwise correlations term can be expressed using the RSQ estimates, which can be, in turn, interpreted as the correlations of asset returns with the systematic factor. As a result, $\sigma_{kl,ij} = \sqrt{T} \cdot cov_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}, \widehat{\rho}_{i\phi}, \widehat{\rho}_{j\phi})$. We denote the covariance matrix of the vector $[\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}, \widehat{\rho}_{i\phi}, \widehat{\rho}_{j\phi}]$ by $\Sigma_{k,l,i,j,\phi}$. The entries of the matrix represent variances and covariances of the pairwise correlation estimates and, thus, can be obtained using Equation (12)³⁷.

We define a two-dimensional function g as follows: $g(\rho_{k\phi}, \rho_{l\phi}, \rho_{i\phi}, \rho_{j\phi}) = [\rho_{k\phi} \cdot \rho_{l\phi}, \rho_{i\phi} \cdot \rho_{j\phi}]^T$. We denote the covariance matrix of $g(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}, \widehat{\rho}_{i\phi}, \widehat{\rho}_{j\phi})$ as Σ_g . It has dimension 2x2 and its off-diagonal term is $cov_T(\widehat{\rho}_{k\phi} \cdot \widehat{\rho}_{l\phi}, \widehat{\rho}_{i\phi} \cdot \widehat{\rho}_{j\phi})$, the quantity of interest.

³⁵ The covariance corresponds to the following term in Equation (12): $cov_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}) = \frac{1}{T} \cdot \sigma_{k\phi,l\phi}$.

³⁶ We use the following gradient notation: $\nabla g(x, y) = \begin{bmatrix} \frac{\partial g}{\partial x}(x, y) \\ \frac{\partial g}{\partial y}(x, y) \end{bmatrix}$.

³⁷ The covariances correspond to the following term in Equation (12): $cov_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}) = \frac{1}{T} \cdot \sigma_{k\phi,l\phi}$.

As in the pairwise correlation case, the delta method allows us to determine the matrix entries. If the delta method assumptions are met, and if T is sufficiently large, we can approximate the matrix $\Sigma_{\mathbf{g}}$ by³⁸ $\nabla \mathbf{g}(\rho_{k\phi}, \rho_{l\phi}, \rho_{i\phi}, \rho_{j\phi})^T \times \Sigma_{k,l,i,j,\phi} \times \nabla \mathbf{g}(\rho_{k\phi}, \rho_{l\phi}, \rho_{i\phi}, \rho_{j\phi})$. The off-diagonal term can be derived straightforwardly as:

$$\begin{aligned}
 & \text{cov}_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}, \widehat{\rho}_{i\phi}, \widehat{\rho}_{j\phi}) = \\
 & = \rho_{l\phi}\rho_{j\phi} \cdot \text{cov}_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{i\phi}) + \rho_{l\phi}\rho_{i\phi} \cdot \text{cov}_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{j\phi}) + \rho_{k\phi}\rho_{j\phi} \cdot \text{cov}_T(\widehat{\rho}_{l\phi}, \widehat{\rho}_{i\phi}) + \rho_{k\phi}\rho_{i\phi} \cdot \text{cov}_T(\widehat{\rho}_{l\phi}, \widehat{\rho}_{j\phi}) \\
 & = \frac{1}{T} \cdot \frac{1}{2} \cdot \rho_{k\phi}\rho_{l\phi}\rho_{i\phi}\rho_{j\phi} \cdot \left\{ (1 - \rho_{k\phi}^2)(1 - \rho_{i\phi}^2) + (1 - \rho_{k\phi}^2)(1 - \rho_{j\phi}^2) + (1 - \rho_{i\phi}^2)(1 - \rho_{j\phi}^2) + \right. \\
 & \quad \left. (1 - \rho_{l\phi}^2)(1 - \rho_{j\phi}^2) \right\} \\
 & = \frac{1}{T} \cdot \frac{1}{2} \cdot \rho_{k\phi}\rho_{l\phi}\rho_{i\phi}\rho_{j\phi} \cdot \left\{ (2 - \rho_{k\phi}^2 - \rho_{i\phi}^2)(2 - \rho_{i\phi}^2 - \rho_{j\phi}^2) \right\} \\
 & = \frac{1}{T} \cdot \frac{1}{2} \cdot \sqrt{RSQ_k RSQ_l RSQ_i RSQ_j} \cdot \left\{ (2 - RSQ_k - RSQ_l)(2 - RSQ_i - RSQ_j) \right\}
 \end{aligned} \tag{25}$$

If we assume a homogeneous portfolio with equal weights and equal RSQ values, we have $\text{cov}_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}, \widehat{\rho}_{i\phi}, \widehat{\rho}_{j\phi}) = \frac{1}{T} \cdot 2 \cdot RSQ^2 \cdot (1 - RSQ)^2$, leading to the convergence, for a large T , which we present in Equation (20): $\text{std}(\widehat{\rho}_{Avg}) \rightarrow \sqrt{\text{cov}_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}, \widehat{\rho}_{i\phi}, \widehat{\rho}_{j\phi})} = \sqrt{\frac{2}{T}} \cdot RSQ \cdot (1 - RSQ)$ as $K \rightarrow \infty$.

³⁸ In our notation, $\nabla \mathbf{g}$ is a matrix of dimension 4×2 : $\nabla \mathbf{g}(\rho_{k\phi}, \rho_{l\phi}, \rho_{i\phi}, \rho_{j\phi}) = \begin{bmatrix} \rho_{l\phi} & 0 \\ \dots & \dots \\ 0 & \rho_{i\phi} \end{bmatrix}$

Appendix B Uncertainty in a Risk Contribution

Section 4.3 discusses the impact of uncertainty in correlations on portfolio-referent, exposure-level risk measures. We now provide more detail on how the uncertainty affects Risk Contribution estimates.

Following the derivations in Moody's Analytics document *Modeling Credit Portfolios*, we can use Equation (22) to express Risk Contribution as:

$$RC_k = \frac{\partial UL}{\partial w_k} = \frac{1}{UL} \cdot \underbrace{\left\{ w_k UL_k^2 + \sum_{l=1}^K w_l LGD_k LGD_l \cdot \{N_2(N^{-1}(PD_k), N^{-1}(PD_l), \rho_{kl}) - PD_k PD_l)\} \right\}}_{\rho_{k,P}^{value}} \quad (26)$$

Assuming we estimate correlations with a statistical error, $\widehat{\rho}_{kl}$, their uncertainty translates into uncertainty of Risk Contribution estimate, \widehat{RC}_k via two channels: \widehat{UL} (the portfolio-level Unexpected Loss) and $\rho_{k,P}^{value}$ (the value correlation between the obligor k and the overall portfolio).

The uncertainty in the portfolio unexpected loss, \widehat{UL} , is driven by uncertainty in all pairwise correlations in the portfolio: $\widehat{\rho}_{kl}$ for all k and l . Using Equation (9) and the results relating to uncertainty in average correlation estimate $\widehat{\rho}_{Avg}$, we can state that, for a large portfolio, the uncertainty in \widehat{UL} is the same for empirical correlations and factor models, for a sufficiently large T .

In contrast, the uncertainty in $\rho_{k,P}^{value}$ depends only on the correlations of the obligor k 's asset return with other obligors in the portfolio: $\widehat{\rho}_{kl}$ for all l . Using the delta method in a similar way to Equation (9), we can show that the uncertainty in $\rho_{k,P}^{value}$ is driven by the uncertainty in $\widehat{\rho}_{Avg,k} = \frac{1}{K} \sum_{l=1}^K w_l \cdot \widehat{\rho}_{kl}$ if T is sufficiently large.

For a large portfolio ($K \rightarrow \infty$), the standard deviation $std(\widehat{\rho}_{Avg,k})$ depends only on covariances of the estimated pairwise correlations, where the pairs have one overlapping index—the index of obligor k : $cov_T(\widehat{\rho}_{kl}, \widehat{\rho}_{kj})$. The reasoning leading to this statement is analogous to the discussion in Equation (15). We can apply Equation (12) to determine $cov_T(\widehat{\rho}_{kl}, \widehat{\rho}_{kj})$, for the case when correlations are estimated as empirical pairwise correlations:

$$cov_T(\widehat{\rho}_{kl}, \widehat{\rho}_{kj}) = \frac{1}{T} \cdot \left\{ \rho_{lj} (1 - \rho_{kl}^2 - \rho_{kj}^2) - \frac{1}{2} \rho_{kl} \rho_{kj} (1 - \rho_{kl}^2 - \rho_{kj}^2 - \rho_{lj}^2) \right\} \quad (27)$$

We can use the delta method, along the lines of Appendix A, to determine $cov_T(\widehat{\rho}_{kl}, \widehat{\rho}_{kj})$ in the case of a single-factor model. Specifically, we have $cov_T(\widehat{\rho}_{kl}, \widehat{\rho}_{kj}) = cov_T(\widehat{\rho}_{k\phi} \cdot \widehat{\rho}_{l\phi}, \widehat{\rho}_{k\phi} \cdot \widehat{\rho}_{j\phi})$ in a single-factor model ($\sqrt{RSQ_k} = \widehat{\rho}_{k\phi}$). Matrix $\Sigma_{k,l,j,\phi}$ is the covariance matrix of $[\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}, \widehat{\rho}_{j\phi}]$ and its entries are provided by Equation (12).

We define $\mathbf{g}(\rho_{k\phi}, \rho_{l\phi}, \rho_{j\phi}) = [\rho_{k\phi} \cdot \rho_{l\phi}, \rho_{k\phi} \cdot \rho_{j\phi}]^T$ and denote the covariance matrix of $\mathbf{g}(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}, \widehat{\rho}_{j\phi})$ as $\Sigma_{\mathbf{g}}$. Its dimension is 2x2 with $cov_T(\widehat{\rho}_{k\phi} \cdot \widehat{\rho}_{l\phi}, \widehat{\rho}_{k\phi} \cdot \widehat{\rho}_{j\phi})$, the quantity of interest, being its off-diagonal term.

If the delta method assumptions are met, and if T is sufficiently large, we can approximate the matrix $\Sigma_{\mathbf{g}}$ by³⁹ $\nabla \mathbf{g}(\rho_{k\phi}, \rho_{l\phi}, \rho_{j\phi})^T \times \Sigma_{k,l,j,\phi} \times \nabla \mathbf{g}(\rho_{k\phi}, \rho_{l\phi}, \rho_{j\phi})$. We can derive the off-diagonal term as:

$$\begin{aligned} cov_T(\widehat{\rho}_{k\phi} \cdot \widehat{\rho}_{l\phi}, \widehat{\rho}_{k\phi} \cdot \widehat{\rho}_{j\phi}) &= \\ &= \rho_{l\phi} \rho_{j\phi} \cdot var_T(\widehat{\rho}_{k\phi}) + \rho_{k\phi} \rho_{j\phi} \cdot cov_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{l\phi}) + \rho_{k\phi} \rho_{l\phi} \cdot cov_T(\widehat{\rho}_{k\phi}, \widehat{\rho}_{j\phi}) + \rho_{k\phi}^2 \cdot cov_T(\widehat{\rho}_{l\phi}, \widehat{\rho}_{j\phi}) \\ &= \frac{1}{T} \cdot \left\{ \rho_{l\phi} \rho_{j\phi} (1 - \rho_{k\phi}^2)^2 + \rho_{k\phi} \rho_{j\phi} \cdot \left[\rho_{kl} (1 - \rho_{k\phi}^2 - \rho_{l\phi}^2) - \frac{1}{2} \rho_{k\phi} \rho_{l\phi} (1 - \rho_{k\phi}^2 - \rho_{l\phi}^2 - \rho_{kl}^2) \right] + \rho_{k\phi} \rho_{l\phi} \right. \\ &\quad \cdot \left[\rho_{kj} (1 - \rho_{k\phi}^2 - \rho_{j\phi}^2) - \frac{1}{2} \rho_{k\phi} \rho_{j\phi} (1 - \rho_{k\phi}^2 - \rho_{j\phi}^2 - \rho_{kj}^2) \right] \\ &\quad \left. + \rho_{k\phi}^2 \left[\rho_{lj} (1 - \rho_{l\phi}^2 - \rho_{j\phi}^2) - \frac{1}{2} \rho_{l\phi} \rho_{j\phi} (1 - \rho_{k\phi}^2 - \rho_{j\phi}^2 - \rho_{lj}^2) \right] \right\} \quad (28) \end{aligned}$$

³⁹ In our notation, $\nabla \mathbf{g}$ is a matrix of dimension 3x2: $\nabla \mathbf{g}(\rho_{k\phi}, \rho_{l\phi}, \rho_{j\phi}) = \begin{bmatrix} \rho_{l\phi} & 0 \\ \dots & \dots \\ 0 & \rho_{j\phi} \end{bmatrix}$

If we assume a large portfolio with uniform weights and uniform RSQ values, calculation of $cov_T(\widehat{\rho}_{kl}, \widehat{\rho}_{kl})$ from either Equation (27) or Equation (28) leads to the same expression: $\frac{1}{T} \cdot \frac{1}{2} \cdot RSQ \cdot \{(2 - RSQ) \cdot (1 - 2 \cdot RSQ^2) + RSQ^3\}$. Thus, $std(\widehat{\rho}_{Avg,k})$, are the same for the pairwise empirical correlation and single-factor cases.

The results above confirm, under simplified circumstances, what we observe in our simulation exercise in Section 4.3:⁴⁰ the uncertainty of Risk Contribution estimate in a large portfolio does not depend on whether the correlations are estimated as empirical pairwise correlations or through a factor model, because the differences between them get "averaged out."

As we mentioned in Section 4.3, the conclusions above would be different for small or heavily concentrated portfolios, where a factor model would exhibit less uncertainty in portfolio-referent, exposure-level risk measures compared to the empirical correlation case.

Although it is not possible to derive similar analytical results for Tail-Risk Contribution, conceptually, the results we discuss for Risk Contributions apply to Tail-Risk Contributions as well.

⁴⁰ Besides individual properties of the two average correlations that we discussed, $\widehat{\rho}_{Avg}$ and $\widehat{\rho}_{Avg,k}$, their covariance matters as well. We could use the same method as in this Appendix to show that their covariance does not depend on whether the correlations are estimated as sample correlations or using a single-factor model.

References

- Board of Governors of the Federal Reserve System, "Guidance on Model Risk Management." *Supervision and Regulatory Letters*, SR 11-7, 2011.
- Broadie, M., "Computing Efficient Frontiers using Estimated Parameters." *Annals of Operations Research: Special Issue on Financial Engineering*, Vol. 45, pp. 21–58, 1993.
- Chopra, V. K. and W. T. Ziemba, "The effect of errors in means, variances, and covariances on optimal portfolio choice." *Journal of Portfolio Management*, Vol. 19, No. 2, pp. 6–11, 1993.
- Gordy, M. B. and E. Heitfield, "Estimating Default Correlations from Short Panels of Credit Rating Performance Data." *Working Paper*, 2002.
- Huang, J., M. Lanfranconi, N. Patel, and L. Pospisil, "Modeling Credit Correlations: An Overview of the Moody's Analytics GCorr Model." Moody's Analytics White Paper, 2012a.
- Huang, J., N. Patel, and L. Pospisil, "Modeling Correlations of Small and Medium-Sized Enterprises." Moody's Analytics Paper, 2012b.
- Levy, A., "An Overview of Modeling Credit Portfolios." Moody's Analytics White Paper, 2008.
- Löffler, G., "The effects of estimation error on measures of portfolio credit risk." *Journal of Banking and Finance*, Vol. 27, No. 8, pp. 1427–1453, 2003.
- Moody's Analytics, "Modeling Credit Portfolios, RiskFrontier Methodology." *Fifth Edition*, 2015.
- Steiger, J. H., "Tests for comparing elements of a correlation matrix." *Psychological Bulletin of the American Psychological Association*, Vol. 87, No. 2, pp. 245–251, 1980.
- Tarashev, N. and H. Zhu, "Modelling and calibration errors in measures of portfolio credit risk," *BIS (Bank for International Settlements) Working Papers*, No. 230, 2007.
- van der Vaart, A. W., "Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)." *Cambridge University Press*, 2000.

© 2018 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "MOODY'S"). All rights reserved.

CREDIT RATINGS ISSUED BY MOODY'S INVESTORS SERVICE, INC. AND ITS RATINGS AFFILIATES ("MIS") ARE MOODY'S CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES, AND CREDIT RATINGS AND RESEARCH PUBLICATIONS PUBLISHED BY MOODY'S ("MOODY'S PUBLICATIONS") MAY INCLUDE MOODY'S CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES. MOODY'S DEFINES CREDIT RISK AS THE RISK THAT AN ENTITY MAY NOT MEET ITS CONTRACTUAL, FINANCIAL OBLIGATIONS AS THEY COME DUE AND ANY ESTIMATED FINANCIAL LOSS IN THE EVENT OF DEFAULT. CREDIT RATINGS DO NOT ADDRESS ANY OTHER RISK, INCLUDING BUT NOT LIMITED TO: LIQUIDITY RISK, MARKET VALUE RISK, OR PRICE VOLATILITY. CREDIT RATINGS AND MOODY'S OPINIONS INCLUDED IN MOODY'S PUBLICATIONS ARE NOT STATEMENTS OF CURRENT OR HISTORICAL FACT. MOODY'S PUBLICATIONS MAY ALSO INCLUDE QUANTITATIVE MODEL-BASED ESTIMATES OF CREDIT RISK AND RELATED OPINIONS OR COMMENTARY PUBLISHED BY MOODY'S ANALYTICS, INC. CREDIT RATINGS AND MOODY'S PUBLICATIONS DO NOT CONSTITUTE OR PROVIDE INVESTMENT OR FINANCIAL ADVICE, AND CREDIT RATINGS AND MOODY'S PUBLICATIONS ARE NOT AND DO NOT PROVIDE RECOMMENDATIONS TO PURCHASE, SELL, OR HOLD PARTICULAR SECURITIES. NEITHER CREDIT RATINGS NOR MOODY'S PUBLICATIONS COMMENT ON THE SUITABILITY OF AN INVESTMENT FOR ANY PARTICULAR INVESTOR. MOODY'S ISSUES ITS CREDIT RATINGS AND PUBLISHES MOODY'S PUBLICATIONS WITH THE EXPECTATION AND UNDERSTANDING THAT EACH INVESTOR WILL, WITH DUE CARE, MAKE ITS OWN STUDY AND EVALUATION OF EACH SECURITY THAT IS UNDER CONSIDERATION FOR PURCHASE, HOLDING, OR SALE.

MOODY'S CREDIT RATINGS AND MOODY'S PUBLICATIONS ARE NOT INTENDED FOR USE BY RETAIL INVESTORS AND IT WOULD BE RECKLESS AND INAPPROPRIATE FOR RETAIL INVESTORS TO USE MOODY'S CREDIT RATINGS OR MOODY'S PUBLICATIONS WHEN MAKING AN INVESTMENT DECISION. IF IN DOUBT YOU SHOULD CONTACT YOUR FINANCIAL OR OTHER PROFESSIONAL ADVISER.

ALL INFORMATION CONTAINED HEREIN IS PROTECTED BY LAW, INCLUDING BUT NOT LIMITED TO, COPYRIGHT LAW, AND NONE OF SUCH INFORMATION MAY BE COPIED OR OTHERWISE REPRODUCED, REPACKAGED, FURTHER TRANSMITTED, TRANSFERRED, DISSEMINATED, REDISTRIBUTED OR RESOLD, OR STORED FOR SUBSEQUENT USE FOR ANY SUCH PURPOSE, IN WHOLE OR IN PART, IN ANY FORM OR MANNER OR BY ANY MEANS WHATSOEVER, BY ANY PERSON WITHOUT MOODY'S PRIOR WRITTEN CONSENT.

All information contained herein is obtained by MOODY'S from sources believed by it to be accurate and reliable. Because of the possibility of human or mechanical error as well as other factors, however, all information contained herein is provided "AS IS" without warranty of any kind. MOODY'S adopts all necessary measures so that the information it uses in assigning a credit rating is of sufficient quality and from sources MOODY'S considers to be reliable including, when appropriate, independent third-party sources. However, MOODY'S is not an auditor and cannot in every instance independently verify or validate information received in the rating process or in preparing the Moody's Publications.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability to any person or entity for any indirect, special, consequential, or incidental losses or damages whatsoever arising from or in connection with the information contained herein or the use of or inability to use any such information, even if MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers is advised in advance of the possibility of such losses or damages, including but not limited to: (a) any loss of present or prospective profits or (b) any loss or damage arising where the relevant financial instrument is not the subject of a particular credit rating assigned by MOODY'S.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability for any direct or compensatory losses or damages caused to any person or entity, including but not limited to by any negligence (but excluding fraud, willful misconduct or any other type of liability that, for the avoidance of doubt, by law cannot be excluded) on the part of, or any contingency within or beyond the control of, MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers, arising from or in connection with the information contained herein or the use of or inability to use any such information.

NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE ACCURACY, TIMELINESS, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OF ANY SUCH RATING OR OTHER OPINION OR INFORMATION IS GIVEN OR MADE BY MOODY'S IN ANY FORM OR MANNER WHATSOEVER.

Moody's Investors Service, Inc., a wholly-owned credit rating agency subsidiary of Moody's Corporation ("MCO"), hereby discloses that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by Moody's Investors Service, Inc. have, prior to assignment of any rating, agreed to pay to Moody's Investors Service, Inc. for appraisal and rating services rendered by it fees ranging from \$1,500 to approximately \$2,500,000. MCO and MIS also maintain policies and procedures to address the independence of MIS's ratings and rating processes. Information regarding certain affiliations that may exist between directors of MCO and rated entities, and between entities who hold ratings from MIS and have also publicly reported to the SEC an ownership interest in MCO of more than 5%, is posted annually at www.moody's.com under the heading "Investor Relations — Corporate Governance — Director and Shareholder Affiliation Policy."

Additional terms for Australia only: Any publication into Australia of this document is pursuant to the Australian Financial Services License of MOODY'S affiliate, Moody's Investors Service Pty Limited ABN 61 003 399 657 AFSL 336969 and/or Moody's Analytics Australia Pty Ltd ABN 94 105 136 972 AFSL 383569 (as applicable). This document is intended to be provided only to "wholesale clients" within the meaning of section 761G of the Corporations Act 2001. By continuing to access this document from within Australia, you represent to MOODY'S that you are, or are accessing the document as a representative of, a "wholesale client" and that neither you nor the entity you represent will directly or indirectly disseminate this document or its contents to "retail clients" within the meaning of section 761G of the Corporations Act 2001. MOODY'S credit rating is an opinion as to the creditworthiness of a debt obligation of the issuer, not on the equity securities of the issuer or any form of security that is available to retail investors. It would be reckless and inappropriate for retail investors to use MOODY'S credit ratings or publications when making an investment decision. If in doubt you should contact your financial or other professional adviser.

Additional terms for Japan only: Moody's Japan K.K. ("MJKK") is a wholly-owned credit rating agency subsidiary of Moody's Group Japan G.K., which is wholly-owned by Moody's Overseas Holdings Inc., a wholly-owned subsidiary of MCO. Moody's SF Japan K.K. ("MSFJ") is a wholly-owned credit rating agency subsidiary of MJKK. MSFJ is not a Nationally Recognized Statistical Rating Organization ("NRSRO"). Therefore, credit ratings assigned by MSFJ are Non-NRSRO Credit Ratings. Non-NRSRO Credit Ratings are assigned by an entity that is not a NRSRO and, consequently, the rated obligation will not qualify for certain types of treatment under U.S. laws. MJKK and MSFJ are credit rating agencies registered with the Japan Financial Services Agency and their registration numbers are FSA Commissioner (Ratings) No. 2 and 3 respectively.

MJKK or MSFJ (as applicable) hereby disclose that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by MJKK or MSFJ (as applicable) have, prior to assignment of any rating, agreed to pay to MJKK or MSFJ (as applicable) for appraisal and rating services rendered by it fees ranging from JP¥200,000 to approximately JP¥350,000,000.

MJKK and MSFJ also maintain policies and procedures to address Japanese regulatory requirements.