Long-Run Stock Return Distributions: Empirical Inference and Uncertainty

Andreas Dzemski Adam Farago Erik Hjalmarsson

Tamas Kiss*

October 14, 2025

Abstract

We analyze empirical estimation of the distribution of total payoffs for stock investments over very long horizons, such as 30 years. Formal results for recently proposed bootstrap estimators are derived and alternative parametric methods are proposed. All estimators should be viewed as inconsistent for longer investment horizons. Valid confidence bands are derived and should be the focus when performing inference. Empirically, confidence bands around long-run distributions are very wide and point estimates must be interpreted with great caution. The scope for distinguishing long-run aggregate return distributions across countries is limited and any significant differences are concentrated to the lowest percentiles.

JEL classification: C58, G1.

Keywords: Estimation uncertainty; Long-run stock returns; Quantile estimation.

*Dzemski, Farago, and Hjalmarsson are with the Department of Economics, University of Gothenburg, P.O. Box 640, SE 405 30 Gothenburg, Sweden. Kiss is with the School of Business, Örebro University, Sweden. Email: andreas.dzemski@economics.gu.se, adam.farago@economics.gu.se, erik.hjalmarsson@economics.gu.se, and tamas.kiss@oru.se. We thank Zhentao Shi as well as seminar participants at the Forecasting Financial Markets Conference, Venice, 2025, and the International Association for Applied Econometrics, Torino, 2025, for helpful comments. Financial support from Marianne and Marcus Wallenberg Foundation (grant number 2019.0117), Jan Wallanders och Tom Hedelius stiftelse samt Tore Browaldhs stiftelse (grant numbers W19-0021 and P19-0117), Vinnova (grant number 2022-00258), and the Skandia Foundation is gratefully acknowledged.

1 Introduction

Many investors face very long horizons for some of their most important investment decisions—such as college, pension, and other life-cycle savings—where holding periods of 30 years or more can be relevant. For investments in risky assets like stocks, the range of possible outcomes over such horizons is extremely large, as the uncertain returns in each period compound over time. In practice, this distribution of outcomes is not known to the investor and needs to be empirically estimated from limited historical data, adding additional uncertainty.

The focus of this study is the formation and sampling uncertainty of empirical estimators of these long-run return distributions. Specifically, we consider different ways of formulating empirical estimates for the payoff of a stock investment over long horizons and how to obtain confidence bounds around these estimates. Empirically, long-run distributions are inherently difficult to pin down, since we at most observe a handful of, say, 30-year returns. Perhaps as a consequence, this question has received relatively little attention in the academic literature. Some recent work has dealt with obtaining empirical point estimates of long-run stock return distributions, but the central question of sampling uncertainty has been left mostly unanswered (see Fama and French, 2018, FF henceforth, and Anarkulova, Cederburg, and O'Doherty, 2022, ACO henceforth).[1] A related recent literature on long-run returns (Bessembinder, 2018, and Farago and Hjalmarsson, 2023ab) highlights that long-run returns tend to be highly asymmetric (positively skewed) and mean forecasts are therefore of limited use. We subsequently focus our analysis on estimates of the entire distribution (i.e., the quantiles).

Our fundamental question is therefore: what is the statistical precision of estimated quantiles of the long-run return distribution? Historical data for many countries begin around 1900, leaving a sample of about 120 annual observations or $n = 1,440$ observations of monthly returns. In terms of long-run returns over, say, $T = 360$ month (30 year) periods,

[1] FF discuss some aspects of precision, but do so in the context of *standardized* results, rescaling the long-run return quantiles by the standard deviation of the long-run returns. The standardization is statistically convenient but recasts the problem in terms of an unknown scale that is not easily interpretable to researchers and practitioners. FF do not present any confidence intervals for their estimates of the actual return quantiles.

this is a short time series that contains only four distinct 30-year returns. *Directly* inferring population characteristics of long-run returns based on such a small sample is clearly fraught with challenges. Instead, it seems more promising to infer such characteristics *indirectly* by compounding the empirical distribution of monthly returns, as suggested in FF and ACO. Both FF and ACO rely on bootstrap procedures, resampling monthly returns. More straightforwardly, one can also think of the indirect approach as simply scaling up the short-run (monthly) features to the appropriate horizon. For example, in a (log) normal model, doubling the horizon doubles the expected return and quadruples the variance, which pins down the distribution at the new longer horizon.

Intuitively, the *indirect* approach offers access to a vastly larger sample than a naïve *direct* estimation ($1,440$ monthly observations rather than four 30-year observations in the above example). Despite this intuitive appeal, it is not clear that the indirect approach will actually yield precise estimates of long-run returns. Compounding estimated monthly returns over a period of T months also compounds the estimation error T-fold. The precision of estimated *monthly* returns, suggested by a large sample of monthly observations, may therefore not translate into precisely estimated long-run returns. We clarify theoretically, and very precisely, how statistical uncertainty about compound returns depends on the relative magnitudes of the compounding horizon and the number of observed monthly returns.

Our theoretical analysis gives a disappointing assessment of the feasibility of obtaining precise point estimates for the distribution of long-run returns. Specifically, we show that the indirect approach faces the same fundamental sample-size constraint as estimation based on actually observed long-run returns (e.g., the observed 30-year returns). This is a feature of the complexity of the estimation task, not the choice of estimator. As our benchmark, we consider a maximum likelihood estimator (MLE) that uses the information contained in the observed monthly returns optimally. By the theory of maximum likelihood, the efficiency of *any* other empirical strategy to infer properties of the long-run returns from monthly returns is bounded by the efficiency of the MLE. If the MLE is inconsistent, or "imprecise" for a given sample size and compounding horizon, then so are all other

estimators, including the bootstrap estimators in FF and ACO or approaches based on long-run returns computed from overlapping time periods.

We find that the rate of convergence for the MLE is identical to the rate of convergence achieved if one were to form T-period returns and directly use these for inference. Second order improvements can be achieved, and these can be important in terms of obtaining better estimates. But "indirect" estimation (including bootstrap procedures) does not avoid the small-sample problem faced in the "direct" approach. For investment horizons and sample sizes that result in only a few unique long-run return observations, one should therefore view all estimators of the long-run distribution as inconsistent, or "imprecise".

Point estimates produced by an imprecise (inconsistent) estimator have to be interpreted with caution. They are the outcome of sampling error as much as the true behavior of long-run returns. Ignoring the sampling error may lead to erroneous conclusions. For example, differences in the estimated return distributions for two countries may appear very large, even though they lie within the margin of sampling error.

We propose a new confidence interval that quantifies the sampling uncertainty about the distribution of long-run returns. Importantly, the confidence interval is valid even if the investment horizon is large, relative to the sample size, and the point estimates are formally inconsistent. This is an advantage of the indirect approach. In contrast, it is not feasible to construct a valid confidence set from the small sample of actually observed T-period returns, without imposing restrictive assumptions about the monthly return distribution. In short, we find that (i) any estimator of the long-run return distribution will inevitably be imprecise, but (ii) we can form valid confidence intervals even when the estimator is not consistent.

Our theoretical analysis also offers new insights about the FF bootstrap. There are no previous formal results on the validity of the FF procedure, but we show that it is consistent under the same rate condition as the MLE.[2] Furthermore, we suggest a new estimator that emulates the positive characteristics of the bootstrap estimator but is

[2]We do not formally analyze the block bootstrap method proposed by ACO, but it is clearly very similar in nature to the regular FF bootstrap. Since the block bootstrap is more general than the FF bootstrap, the ACO block bootstrap will require at least the same conditions for consistency as the FF bootstrap.

computationally trivial and easy to implement. Indeed, an important and useful takeaway from our analysis is that the FF bootstrap can be viewed as a skewness-corrected version of the MLE under log-normality.

Precision of the estimators can be improved by using panel data, rather than a single time series. However, even with a large cross section, the gains are limited by the cross-sectional correlation of asset returns. We illustrate this using a simple return model with a single market factor. If the common factor explains, say, about a third of the overall variation in returns, access to a large panel will be equivalent to multiplying the time-series dimension by about 3. That is, a panel with, say, 30 countries and 1,000 monthly observations for each country, would yield estimates of the same precision as a single time-series with 3,000 observations. This is an important improvement over the pure time-series case, but much less than one might expect from the vast increase in the nominal sample size offered by the panel. The availability of panel data also raises the question of whether the return series share a common underlying distribution and we develop a formal test of equality of long-run return distributions.

In the empirical analysis, we revisit the question of the long-run distribution of global stock returns, analyzed by ACO. We use the Dimson, Marsh, and Staunton (DMS) data set, which provides annual stock return observations for 21 countries from 1900 to 2020. Like ACO, we focus on real returns measured in local currency.

Our main empirical contribution is to quantify estimation uncertainty, both for the long-run return distributions themselves and for cross-country differences in these distributions. Our confidence intervals provide statistical error bands around our estimates of the global and country-level long-run return distributions. In addition, our formal test evaluates the statistical significance of apparent differences in the long-run return distributions across countries.

It is sometimes posited that the U.S., in particular, is a "lucky survivor" and that its historically strong stock market performance might not be representative of other countries (e.g., Goetzmann and Jorion, 1999, and van Binsbergen et al., 2023). Here we study whether long-run payoffs on global stock markets are supportive of this conjecture

when we account for sampling uncertainty.

We estimate quantiles for the long-run global stock returns by pooling all the data. The resulting global estimates are compared to individual country-level estimates. When viewed without accounting for estimation uncertainty, the global and U.S. long-run return distributions look strikingly different. For example, the estimate of the 5^{th} percentile of net 30-year global real stock returns is -55% and the corresponding figure for the U.S. is $+14\%$, lending support to the notion that the U.S. is different.[3] However, the confidence bands around these estimates are wide and overlap substantially: the 90% confidence interval for the global 5^{th} percentile stretches from -80% to -3%, and the confidence interval for the U.S. 5^{th} percentile stretches between -57% and 204%. The estimate of the 50^{th} percentile (i.e., the median) of net 30-year global real stock returns is 307% and for the U.S. 597%. But the corresponding 90% confidence intervals are again extremely wide: for the global median it ranges from 104% to 709%, and for the U.S. median from 192% to 1566%. Our formal test does reject equality for the 5^{th} percentile (p-value = 0.03), but fails to reject equality for higher percentiles such as the median (p-value = 0.17).

We extend the analysis beyond the U.S. to the other 20 countries in our sample. Four main patterns emerge. First, when testing each country in turn, we detect several countries that differ significantly from the global return distribution. Second, the U.S. is not among those with the most pronounced differences. Third, significant differences are found almost exclusively in the lower tail (our analysis focuses on the 5^{th} percentile). Fourth, once we account for the multiple testing problem inherent in comparing many countries simultaneously, nearly all evidence of significant differences vanishes, and the U.S. does not stand out as distinct from the global distribution.

The exact statistical significance that one attaches to the U.S. evidence depends on the extent to which one is willing to *ex ante* assign the U.S. a special role and thus (implicitly) ignore the multiple testing issue. At the same time, multiple-testing corrections are unavoidably conservative, so the absence of significance cannot be taken as evidence of the absence of differences.

[3]We consider three different estimators in our empirical analysis, all yielding very similar results. The figures reported here correspond to the skewness-corrected MLE.

Regardless of the precise interpretation, our analysis highlights the inherent difficulty of distinguishing long-run return properties across countries. This lack of statistical power is rooted in a fundamental scarcity of data: even the longest available time series yield only a handful of independent 30-year returns. The small-sample nature of the problem underscores the importance of reporting confidence intervals alongside point estimates. Nearly all aspects of the long-run distributions are measured with considerable imprecision, and without explicit error bounds and formal tests, it is easy to place undue weight on the point estimates and the large apparent differences between them.

2 The big picture

Before delving into the formal theory, it is useful to first provide a simplified bird's-eye view of the problem we are trying to address. The basic setup is a follows. We have n return observations at some moderately high frequency. In our main example, we observe monthly returns over 120 years, such that $n = 1,440$. Based on this sample, we wish to perform inference on the total (compound) return over some investment horizon T.

If T is small relative to n, one could simply aggregate the original monthly data into n/T (non-overlapping) observations of T-period returns. If $T = 12$, the monthly sample with $n = 1,440$ returns would aggregate to a sample of $n/T = 120$ annual returns. In these aggregated data, the frequency of observation (annual) coincides with the return horizon that we want to study, and standard statistical procedures can be applied. We refer to this as "direct" inference.

When T is large relative to n, such that the number of distinct T-period observations is small, direct inference seems less attractive. With $n = 1,440$ and $T = 360$, there are only four 30-year observations. Estimation based on such a small sample will be imprecise and, importantly, diagnostic tools such as confidence intervals are not guaranteed to be valid (unless very strict assumptions are imposed).[4]

[4]An alternative to a straightforward aggregation to T-period returns, would be to form *overlapping* T-period observations. This may in some cases increase the efficiency of the estimation, relative to the case with non-overlapping observations. However, the MLE that we subsequently consider is the most efficient estimator given the initial monthly observations, and will therefore also be more efficient than any estimator based on the overlapping T-period observations.

With T large, it therefore seems more promising to infer the long-run distribution "indirectly" from the observed monthly returns. In this vein, FF and ACO resample monthly returns to create an arbitrarily large number of bootstrapped T-period returns. Alternatively, one can estimate the properties of the one-period returns and then extrapolate to the long-run returns. In both cases, estimation is based on a sample that is seemingly much larger than the sample of observed long-run (T-period) returns.

The key question that we answer in our formal analysis is to what extent such "indirect" methods actually escape the (very) small sample-size problem that one faces in the "direct" approach. As discussed in detail in the following section, the disappointing answer is that indirect methods are inevitably subject to the same fundamental sample-size problems as the direct approach, and that the *overall rate of convergence* cannot be improved upon. This is not to say that the direct estimation cannot be improved upon; in practice, second order improvements to an estimator can be important. But the first-order precision of any estimator depends upon the relative magnitudes of T and n, which cannot be altered for a given sample size and return horizon. On a more positive note, we show that confidence intervals based on indirect methods are valid even if the number of observed long-run returns is small and without imposing strong parametric assumptions. This allows us to empirically assess the (im-)precision of estimated long-run return distributions, which is crucial in a setting where the point estimates are subject to great sampling uncertainty.

To provide some intuition for our results, we now consider a simplified example. Suppose that one-period log returns are independent and identically distributed (i.i.d.) normal (gross returns are log-normal) with unknown mean μ_y and known variance σ_y^2. In this case, the log T-period returns are i.i.d. normal with mean $\mu_Y = T\mu_y$ and variance $T\sigma_y^2$. Assuming that n/T is an integer, denote the original monthly log returns as $y_1, ..., y_n$, and the corresponding log T-period returns as $Y_1, Y_2, ..., Y_{n/T}$, where $Y_1 = y_1 + ... + y_T, Y_2 = y_{T+1} + ... + y_{2T}$, and so forth. A direct way to obtain an estimate of the unknown mean μ_Y is to take the average of the observed T-period returns. An indirect way is to first

estimate the expected single-period log return μ_y, using

$$\hat{\mu}_y = \frac{1}{n}\sum_{t=1}^{n} y_t,$$

and then scale up this estimate to the long-run horizon by multiplying by T. But both approaches yield the same estimate of μ_Y, since

$$\hat{\mu}_Y = \frac{1}{n/T}\sum_{s=1}^{n/T} Y_s = \frac{T}{n}\sum_{s=1}^{n/T}\left(\sum_{q=1}^{T} y_{(s-1)\times T+q}\right) = \frac{T}{n}\sum_{t=1}^{n} y_t = T \times \hat{\mu}_y. \tag{1}$$

That is, the direct estimator of the mean log T-period return ($\hat{\mu}_Y$) is identical to the mean estimator formed from the original monthly returns, multiplied by the return horizon T ($T \times \hat{\mu}_y$). In this simplified normal case, with a known variance, the estimator of the *distribution* of the long-run T-period returns (which is a function of the mean and variance alone) is therefore identical irrespective of whether one performs direct inference on the T-period returns or whether one indirectly infers the long-run distribution from the monthly returns.

To get some sense of the precision of $\hat{\mu}_Y$, note that

$$\text{var}\left(\hat{\mu}_Y\right) = \frac{1}{n^2/T^2}\sum_{s=1}^{n/T}\text{var}\left(Y_s\right) = \frac{T^2}{n}\sigma_y^2, \tag{2}$$

since the Y_s are independent with variance $T\sigma_y^2$. This implies that the standard error of $\hat{\mu}_Y$ is proportional to $T/\sqrt{n}$. In order for $\hat{\mu}_Y$ to be "precise", $T/\sqrt{n}$ needs to be "small".

In our formal theoretical analysis, we show that this rate condition—trivially derived here, in our simplified example—is in fact a fundamental feature of the estimation problem, restricting direct as well as indirect estimation approaches. The rate condition does not rely on the normality assumption, but holds for general distributions of the one-period returns. The obvious small-sample problem in the direct estimation approach is therefore more binding than is at first apparent, even though the direct and indirect approaches will not in general be equivalent. In the following section, we formalize these claims.

3 Estimating the distribution of long-run returns

In this section, we formulate the estimation problem and derive theoretical results for different inferential approaches. In particular, we show under what conditions (i) consistency of an estimator can be achieved and (ii) valid confidence intervals can be obtained. We focus on asymptotic (large sample) properties, since finite sample results are typically only obtainable under very restrictive assumptions.

3.1 Asymptotic framework

We consider settings where both the investment horizon T and the number of observed one-period returns n are large. A key feature of our analysis is that we do not restrict the number of long-run returns, n/T, nor the ratio $T/\sqrt{n}$ introduced previously. The number of long-run return observations, n/T, is therefore allowed to be small even as n grows large, which mirrors the practical situation faced by empirical researchers observing a large number of one-period returns but only a small number of distinct long-run returns. Formally, we consider the asymptotic limit as the size n of the estimation sample grows large and allow the investment horizon, T, to depend on the sample size, i.e., $T = T(n)$.

A similar asymptotic framework is used in the literature on long-horizon forecasting regressions, where the forecasting horizon is sometimes treated as a finite fraction of the sample size (e.g., Richardson and Stock, 1989, Valkanov, 2003). However, our approach is more general in that we do not assume T and n to grow at the same rate. Rather, we allow T to grow at any rate, but find restrictions on this rate under which consistency and inferential validity hold.

3.2 Notation and setup

Let x_t represent the *one-period gross return* from $t-1$ to t. For most of our analysis, we take one period to be one month, and x_t to represent monthly returns. The primary exception is in the empirical section, where we work with annual returns data.

The long-run compound return, X_T, accruing from an investment at time 0 held until

time T, is defined as

$$X_T \equiv x_1 \times x_2 \times ... \times x_T \ . \tag{3}$$

Our focus is on the case when T is "large" and we use $T = 120$ months (10 years) and $T = 360$ months (30 years) in our implementations.

Our goal is to conduct statistical inference on the quantiles of X_T. We denote the τ-quantile of X_T by $Q_{X_T}(\tau)$. Setting $\tau = 0.5$ gives the median long-run return and small τ-quantiles correspond to the value-at-risk of the long-run investment. For example, the 0.05-quantile gives the minimal long-run return after excluding the worst outcomes with a combined probability of 5%. Quantiles are measured in gross dollar-returns on a dollar invested in period zero. For example, a gross return of 2 implies a 100% net return and a gross return of 0.5 implies a -50% net return. For readability, we will often refer to *percentiles* (e.g. 50^{th}) rather than quantiles (0.5).

The gross compound returns X_T are our primary interest, since these represent the returns actually accruing to investors. Analytically, it is convenient to work with log-transformed returns, as it allows us to appeal to large-sample approximations of sums, such as the central limit theorem (CLT). The log-transformed one-period returns are denoted by $y_t \equiv \log(x_t)$, and the long-run log returns by

$$Y_T \equiv \log(X_T) = y_1 + ... + y_T. \tag{4}$$

Let $Q_{Y_T}(\tau)$ denote the τ-quantile of Y_T. The quantiles of the gross returns are completely characterized by the quantiles of the log-transformed returns via the equality

$$Q_{X_T}(\tau) = \exp\left(Q_{Y_T}(\tau)\right). \tag{5}$$

Based on this equivalence, we can construct estimators and confidence intervals for Q_{X_T} from estimators and confidence intervals for Q_{Y_T}. This approach does not extend to other features of the return distribution. For example, for the expected compound returns, $\mathbb{E}[X_T] \neq \exp(\mathbb{E}[Y_T])$.

Per equation (3), the joint distribution of single-period returns x_t completely pins

down the distribution of the compound return X_T. If the x_ts are i.i.d., the marginal distribution of x_t completely specifies the return generating process. If the returns are dependent over time (or exhibit heterogeneity), these aspects also need description. We start by considering the i.i.d. setting. It admits a clear and intuitively interpretable characterization of the role of sampling error. In Section 3.5, we discuss some extensions to serially correlated returns.

3.3 Log-normal returns

To illustrate the difficulty of statistical inference on long-run returns in a simple setting, we first consider a parametric specification of the return process. Specifically, we assume that the single-period returns x_t are i.i.d. log-normal. Equivalently, we assume that the single-period log returns are i.i.d. normal, i.e.,

$$y_t \sim N\left(\mu_y, \sigma_y^2\right), \tag{6}$$

for unknown parameters μ_y and $\sigma_y^2 > 0$, with y_t independent across time. We interchangeably refer to this parametric setting as the normal or log-normal case.

The long-run log returns are normally distributed as

$$Y_T \sim N\left(T\mu_y, T\sigma_y^2\right), \tag{7}$$

and the τ-quantile of Y_T is given by

$$Q_{Y_T}(\tau) = T\mu_y + \sqrt{T}\sigma_y \Phi^{-1}(\tau), \tag{8}$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative normal distribution function. Plugging this equality into (5) yields an analytic expression for the quantiles of the gross long-run return, X_T.

For a parametric return process, standard optimality arguments imply that the most efficient estimator of the long-run return distribution is the maximum likelihood estimator (MLE). This means that the MLE achieves the highest precision of any possible estimator

and makes it a valuable benchmark to assess the feasibility of consistent estimation of the long-run returns.

The MLE of the τ-quantile of the logged long-run return is given by

$$\widehat{Q}^{\mathrm{ML}}_{Y_T}(\tau) = T\hat{\mu}_y + \sqrt{T}\hat{\sigma}_y \Phi^{-1}(\tau), \tag{9}$$

where

$$\hat{\mu}_y = \frac{1}{n}\sum_{t=1}^{n} y_t \quad \text{and} \quad \hat{\sigma}_y^2 = \frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{\mu}_y)^2 . \tag{10}$$

The MLE of the τ-quantile of the gross long-run return is given by

$$\widehat{Q}^{\mathrm{ML}}_{X_T}(\tau) = \exp\left(\widehat{Q}^{\mathrm{ML}}_{Y_T}(\tau)\right). \tag{11}$$

The following proposition characterizes consistency of the ML estimator.

Proposition 1 (**Consistent estimation in normal model**). *Suppose that single-period log returns, y_t, are i.i.d. normal or, equivalently, that single-period gross-returns, x_t, are i.i.d. log-normal. Then the ML quantile estimator $\widehat{Q}^{ML}_{Y_T}(\tau)$ is consistent for $Q_{Y_T}(\tau)$, such that*

$$\widehat{Q}^{ML}_{Y_T}(\tau) - Q_{Y_T}(\tau) \xrightarrow{p} 0 \quad \textit{as } n \to \infty,$$

if and only if $T(n)/\sqrt{n} \to 0$.

Proof. See Online Appendix. □

This result ties consistency of the MLE to the fundamental ratio $T(n)/\sqrt{n}$ that we already motivated in Section 2. The MLE for the log-return distribution is consistent if and only if the ratio $T(n)/\sqrt{n}$ vanishes as n grows large.[5] This is an asymptotic condition that can be interpreted such that estimates are precise if this ratio is "sufficiently small". In

[5]Since $T = T(n)$ changes with n as $n \to \infty$, consistency is based on the distance between an estimator and a potentially drifting sequence of target parameters. This is different from the standard asymptotic framework where the target parameter is fixed.

our leading example with $n = 1,440$ and $T = 360$, $T/\sqrt{n} \approx 9$. In subsequent simulations, we show that under reasonable parameterizations of the the return process, this value is not "sufficiently small" and precise estimation is not feasible.[6]

Our parameter of interest is the gross return, not the log return. In Proposition A1 in the Online Appendix, we show that the conditions for consistent estimation of the gross return are even more restrictive than for the log return.[7]

Proposition 1 has implications beyond the log-normal model. Given the optimality properties of ML estimation, inconsistency of the MLE suggests inconsistency of all other estimators of the normal model. Furthermore, imposing log-normality reduces the complexity of the estimation task. If consistency cannot be achieved in the log-normal model, then it cannot be achieved in more general models with less restrictive distributional assumptions. Therefore, Proposition 1 suggests a general impossibility result: if its necessary conditions are not met, then consistent estimation of the long-run return distribution is not possible. This is an intrinsic property of the complexity of the estimation task and cannot be overcome by developing more sophisticated methods.

As an alternative to point estimation, we now consider interval estimation and provide confidence intervals that are valid under much weaker conditions than those required for consistent estimation. That is, even though consistent estimation might not be feasible, valid confidence intervals are usually attainable.[8] The confidence intervals characterize the quantiles of the long-run returns up to an error margin that reflects fundamental statistical uncertainty.

Let the lower and upper bounds of a $(1-\alpha)$-confidence interval for the τ-quantile

[6] Our confidence intervals allow us to link the ratio $T/\sqrt{n}$ to a finite-sample error margin. This is discussed below and makes the meaning of "sufficiently small" more precise.

[7] If T was treated as fixed, consistency of $Q_{X_T}(\tau)$ would follow from consistency of $Q_{Y_T}(\tau)$ via Slutsky's lemma. This is not true for the large T case. Given that our primary focus is on valid confidence intervals, where validity for $Q_{Y_T}(\tau)$ immediately implies validity for $Q_{X_T}(\tau)$, we relegate the consistency result for gross returns to the Online Appendix.

[8] Intuitively, our confidence intervals provide valid error bounds, but these bounds are not guaranteed to shrink to zero as the sample size grows. Valid confidence intervals do not require the estimator to be consistent, as illustrated by finite-sample inference for parametric models.

$Q_{Y_T}(\tau)$ be given by

$$\widehat{Q}^{\ell,\mathrm{ML}}_{Y_T}(\tau;\alpha) = \widehat{Q}^{\mathrm{ML}}_{Y_T}(\tau) - \frac{T}{\sqrt{n}}\hat{\sigma}_y \Phi^{-1}\left(1-\alpha/2\right)\psi_T(\tau), \tag{12}$$

$$\widehat{Q}^{u,\mathrm{ML}}_{Y_T}(\tau;\alpha) = \widehat{Q}^{\mathrm{ML}}_{Y_T}(\tau) + \frac{T}{\sqrt{n}}\hat{\sigma}_y \Phi^{-1}\left(1-\alpha/2\right)\psi_T(\tau), \tag{13}$$

where

$$\psi^2_T(\tau) = 1 + \frac{1}{2T}\left(\Phi^{-1}(\tau)\right)^2. \tag{14}$$

The corresponding lower and upper bounds of the confidence interval for the τ-quantiles of the gross returns, $Q_{X_T}(\tau)$, are obtained via equation (5) and are given by

$$\widehat{Q}^{\ell,\mathrm{ML}}_{X_T}(\tau;\alpha) = \exp\left(\widehat{Q}^{\ell,\mathrm{ML}}_{Y_T}(\tau;\alpha)\right), \tag{15}$$

$$\widehat{Q}^{u,\mathrm{ML}}_{X_T}(\tau;\alpha) = \exp\left(\widehat{Q}^{u,\mathrm{ML}}_{Y_T}(\tau;\alpha)\right). \tag{16}$$

The two terms in ψ^2_T can be interpreted as multipliers of different sources of uncertainty. The first term emerges from uncertainty about the mean μ_y, whereas the second term is due to uncertainty about the variance σ^2_y. As T grows, the relative contribution of the second term decreases, leaving the mean as the dominant source of uncertainty. An important implication is that increasing the sampling frequency of the data used for estimating μ_y and σ^2_y (e.g., from annual to monthly or from monthly to daily) will at most have a second order effect on the precision of the quantile estimator. It is well known that increasing the sampling frequency will increase the precision of the variance estimator, but not the mean estimator (e.g., Merton, 1980). In simulations reported in the Online Appendix, we find that the quantile estimator has virtually identical precision regardless of whether one uses daily, monthly, or annual data.[9]

[9]There is clearly a limit to how low a sampling frequency that one can use, before affecting the precision of the quantile estimator. In the extreme case, one only samples T-period returns, which results in what we refer to as the "direct" estimation approach. In our example with 120 years of data and a return horizon of 30 years, this leaves a mere four observations. In this case, the second order effects become important, because the estimator of the variance is now very imprecise, as is also illustrated by simulation results in the Online Appendix. Thus, while the direct approach formally has the same rate of convergence as the indirect approaches, the latter will still perform better in finite samples.

The following proposition establishes that our confidence intervals are valid in the sense that they cover the true quantile with probability approaching the nominal coverage probability $1-\alpha$.

Proposition 2 (**Confidence intervals for normal model**). *Suppose that single-period log returns, y_t, are i.i.d. normal or, equivalently, that single-period gross-returns, x_t, are i.i.d. log-normal. Then, for any $\tau \in (0,1)$ and any sequence $T = T(n)$,*

$$P\left(\widehat{Q}_{Y_T}^{\ell,ML}(\tau;\alpha) \leq Q_{Y_T}(\tau) \leq \widehat{Q}_{Y_T}^{u,ML}(\tau;\alpha)\right) \to 1-\alpha, \tag{17}$$

$$P\left(\widehat{Q}_{X_T}^{\ell,ML}(\tau;\alpha) \leq Q_{X_T}(\tau) \leq \widehat{Q}_{X_T}^{u,ML}(\tau;\alpha)\right) \to 1-\alpha, \tag{18}$$

as $n \to \infty$.

Proof. See Online Appendix. □

The coverage guarantee in Proposition 2 requires no restrictions on the investment horizon T. This mirrors results in the long-run predictability literature, where testing may be possible even when consistent estimation is not (e.g., Valkanov, 2003, and Hjalmarsson and Kiss, 2022).

The investment horizon affects the width of the confidence intervals, but not their validity. The width is related to the familar ratio $T/\sqrt{n}$, allowing us to give a more precise interpretation of the condition that this ratio must be small for the MLE to be precise (see Proposition 1).

3.4 Non-normal returns

We now abandon the assumption of log-normal returns and discuss nonparametric estimation for settings where the marginal distribution of the one-period return x_t is unknown. We maintain the assumption of independence over time.[10]

[10] Our theoretical results are derived under an i.i.d. assumption. Our arguments exploit independence, whereas the assumption of identical distribution is less crucial. A relaxation of the "identical" condition seems possible, but would require cumbersome notation, additional technical conditions, and less transparent theoretical results and arguments. We do not pursue such extensions here.

3.4.1 A correction for non-normality

For a long investment horizon T, central limit theory suggests that the *average* monthly log return approximately follows a normal distribution. This in turn suggests that the MLE estimators $\widehat{Q}^{\text{ML}}_{Y_T}(\tau)$ and $\widehat{Q}^{\text{ML}}_{X_T}(\tau)$ may produce informative estimates even if the assumption of log-normal one-period returns does not hold; FF make a similar point in their study.

We formalize this intuition and show (see Lemma A4 in the Online Appendix) that under i.i.d. but non-normal returns,

$$Q_{Y_T}(\tau) = T\mu_y + \sqrt{T}\sigma_y \Phi^{-1}(\tau) + \frac{1}{6}\sigma_y\gamma_y\left(\left(\Phi^{-1}(\tau)\right)^2 - 1\right) + O\left(T^{-1/2}\right), \tag{19}$$

where

$$\gamma_y = \frac{\mathbb{E}\left[(y_t - \mu_y)^3\right]}{\sigma_y^3}, \tag{20}$$

is the skewness of y_t. Comparing equations (19) and (8) reveals that the quantiles of the long-run log return can be written as the sum of the corresponding quantiles under normality, a skewness correction, and a higher-order approximation error. In particular, for non-skewed log returns and an investment horizon $T \to \infty$, the MLE for the (possibly misspecified) log-normal model will be consistent under the conditions of Proposition 1. For skewed returns, the MLE is biased even if the investment horizon T is large.

Equation (19) suggests a skewness-corrected ML estimator given by

$$\widehat{Q}^{\text{ML-skew}}_{Y_T}(\tau) = T\hat{\mu}_y + \sqrt{T}\hat{\sigma}_y \Phi^{-1}(\tau) + \frac{1}{6}\hat{\sigma}_y\hat{\gamma}_y\left(\left(\Phi^{-1}(\tau)\right)^2 - 1\right), \tag{21}$$

with

$$\hat{\gamma}_y = \frac{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{\mu}_y)^3}{\hat{\sigma}_y^3}. \tag{22}$$

An estimator of the gross returns is given by $\widehat{Q}^{\text{ML-skew}}_{X_T}(\tau) = \exp\left(\widehat{Q}^{\text{ML-skew}}_{Y_T}(\tau)\right)$.

Below, we show that $\widehat{Q}^{\text{ML-skew}}_{Y_T}(\tau)$ is consistent for $Q_{Y_T}(\tau)$ under the same rate condition

as stated in Proposition 1. Before stating the formal result, we first discuss an alternative nonparametric estimator.

3.4.2 The FF bootstrap

For the case with i.i.d., but not necessarily log-normal returns, FF develop the following nonparametric bootstrap procedure to estimate the distribution of the long-run log return.

1. Sample, with replacement, T observations from $\{y_t\}_{t=1}^n$. This gives a bootstrapped return series $y_1^*, y_2^*, ..., y_T^*$, from which a bootstrapped long-run return $Y_T^* = y_1^* + y_2^* + \cdots + y_T^*$ is calculated.

2. Repeat Step 1 B times, where B is a large number (FF use $B = 100,000$ and we use the same value in our implementation). This generates a collection of bootstrapped long-run returns, $\left\{Y_{T,b}^*\right\}_{b=1}^B$.

3. The cumulative distribution function (cdf) of Y_T is estimated by the empirical cdf of the bootstrapped returns, $\left\{Y_{T,b}^*\right\}_{b=1}^B$,

$$\hat{F}_{Y_T}^{\text{boot}}(z) = \frac{1}{B}\sum_{j=1}^{B} \mathbf{1}\left\{Y_{T,b}^* \leq z\right\}, \tag{23}$$

and the FF bootstrap estimate of the τ-quantile of Y_T is given by

$$\widehat{Q}_{Y_T}^{\text{boot}}(\tau) = \inf\left\{z : \hat{F}_{Y_T}^{\text{boot}}(z) \geq \tau\right\}. \tag{24}$$

A corresponding bootstrap estimator for the gross return is given by $\widehat{Q}_{X_T}^{\text{boot}}(\tau) = \exp\left(\widehat{Q}_{Y_T}^{\text{boot}}(\tau)\right)$.[11]

The FF bootstrap generates a large sample of synthetic long-run returns by resampling from the empirical distribution of one-period returns. This is an intuitively appealing

[11] One can equivalently sample the gross one-period returns in Step 1 and calculate gross long-run returns $X_T^* = x_1^* \times x_2^* \times \cdots \times x_T^*$. $\hat{F}_{X_T}^{\text{boot}}(z)$ and $\widehat{Q}_{X_T}^{\text{boot}}(\tau)$ can then be directly calculated from the bootstrapped gross returns $\left\{X_{T,b}^*\right\}_{b=1}^B$, analogously to Step 3. This approach is more natural if one is interested in the distribution of the gross returns. We present the bootstrap in terms of the log returns to conform with the other estimation methods presented here, which all start with the log returns.

remedy to the problem of observing only a small number of actual long-run returns. The number of synthetic T-period returns is limited only by computing power.

The FF bootstrap is completely nonparametric in the sense that it makes no assumptions about the marginal distribution of the one-period return x_t. However, it crucially depends on the assumption of independent returns to ensure that single-period returns can be re-shuffled in any order without affecting the properties of the compound returns. ACO relax this restriction, as discussed further in Section 3.5.

The synthetic returns recycle information from limited historical data and are therefore not a replacement for a large sample of actual long-run returns. FF document this in a simulation study where they compare estimation based on B re-sampled (bootstrapped) long-run returns to estimation based on B observations of actual long-run returns. They find the latter estimator to have much lower variance than the bootstrap estimator. This suggests that re-sampling cannot overcome the inherent small sample problem that arises with long investment horizons.

FF do not formally validate their bootstrap procedure or provide any results on its theoretical properties. In the next section, we provide a sharp characterization of the validity of the FF bootstrap, which clearly identifies the role of the relative magnitude of T to n. This characterization is not obvious from standard bootstrap theory, since the bootstrapped returns depend on $T = T(n)$, with possibly $T(n) \to \infty$.

3.4.3 Inference based on the skewness-corrected MLE and the FF bootstrap

The following proposition characterizes consistency of the skewness-corrected MLE and the FF bootstrap estimator.

Proposition 3 (**Consistent estimation in nonparametric model**). *Suppose that the single-period log returns y_t are i.i.d., have a continuous density, and have eight moments. Then $\widehat{Q}_{Y_T}^{ML\text{-}skew}(\tau)$ and $\hat{Q}_{Y_T}^{boot}(\tau)$ are consistent for $Q_{Y_T}(\tau)$ if and only if $T(n)/\sqrt{n} \to 0$ as $n \to \infty$.*

Proof. See Online Appendix. □

The rate condition, in terms of the crucial ratio $T/\sqrt{n}$, is the same as for ML estimation of the parametric log-normal model.

The proof of Proposition 3 proceeds by showing that the skewness-corrected MLE and the FF bootstrap estimator are first-order asymptotically equivalent. This equivalence leads to the neat interpretation of the FF bootstrap estimator as an alternative way of implementing the skewness-corrected MLE.

Given the strict rate restrictions required for consistent estimation—which are arguably often not satisfied empirically—we focus on obtaining valid confidence intervals. We construct confidence intervals for the log returns, which can then be transformed to confidence intervals for the gross returns via equation (5).

For sequences $T = T(n) \to \infty$, the skewness-adjustment in equation (21) is of higher order than the other terms. That is, it does not first-order contribute to sampling uncertainty. Therefore, confidence intervals can be constructed by re-centering the confidence interval for the MLE on the skewness-corrected ML estimate or the FF bootstrap estimate. This is formalized in the following proposition.

Proposition 4 (**Confidence intervals for nonparametric model**). *Suppose that the single-period log returns y_t are i.i.d., have a continuous density, and have eight moments. Let $\widehat{Q}_{Y_T}(\tau)$ denote either the FF bootstrap estimator or the skewness-corrected MLE and let*

$$\widehat{Q}^{\ell}_{Y_T}(\tau) = \widehat{Q}^{\ell,ML}_{Y_T}(\tau) + \widehat{Q}_{Y_T}(\tau) - \widehat{Q}^{ML}_{Y_T}(\tau), \tag{25}$$

$$\widehat{Q}^{u}_{Y_T}(\tau) = \widehat{Q}^{u,ML}_{Y_T}(\tau) + \widehat{Q}_{Y_T}(\tau) - \widehat{Q}^{ML}_{Y_T}(\tau). \tag{26}$$

Then, for any $\tau \in (0,1)$ and any sequence $T = T(n)$ such that $T(n)^{3/2}/\sqrt{n} \to \infty$,

$$P\left(\widehat{Q}^{\ell}_{Y_T}(\tau;\alpha) \leq Q_{Y_T}(\tau) \leq \widehat{Q}^{u}_{Y_T}(\tau;\alpha)\right) \to 1-\alpha, \tag{27}$$

as $n \to \infty$.

Proof. See Online Appendix. □

A confidence interval for the gross return can again be constructed by exponentiating the lower and upper bounds of the confidence interval for the log returns. That is, its lower and upper bounds are given by,

$$\widehat{Q}^{\ell}_{X_T}(\tau;\alpha) = \exp\left(\widehat{Q}^{\ell}_{Y_T}(\tau;\alpha)\right), \tag{28}$$

$$\widehat{Q}^{u}_{X_T}(\tau;\alpha) = \exp\left(\widehat{Q}^{u}_{Y_T}(\tau;\alpha)\right), \tag{29}$$

and the resulting confidence interval is valid under the conditions of Proposition 4.

Proposition 4 imposes the rate condition $T(n)^{3/2}/\sqrt{n} \to \infty$. This condition ensures that T is large enough for the pull of central limit theory to justify the approximation in equation (19). A sufficient condition for the rate condition to be satisfied is that $T(n)$ and n are of the same order of magnitude, which arguably holds in our examples with compounding horizons in excess of 10 years and sample periods spanning around 100 years.

3.5 Non-i.i.d. returns

The inferential methods described above are derived under the assumption of i.i.d. one-period returns. As mentioned in footnote 10, violations of the "identical" part of the i.i.d. assumption (e.g., time-varying variance) are unlikely to affect the validity of the inferential approach. Dependence, and in particular serial correlation in returns, might impact inference. However, serial dependence in stock returns is typically weak. Given the bias-variance trade-off introduced by more general estimators, it is therefore not obvious that trying to accommodate serial correlation will necessarily lead to better estimates.

In the simulations in Section 5, we show results for a stochastic volatility model, which is a martingale difference sequence rather than an i.i.d. sequence, and find that the methods derived in the i.i.d. setting works well in this case. We also study the proposed estimators under a long-term mean reverting return process. For empirically reasonable levels of serial dependence, inference is only marginally impacted and the proposed confidence intervals retain very good coverage rates. Additional simulations in the Online Appendix show that the ACO block bootstrap estimator, designed to control for serial correlation in returns,

can in some cases reduce bias, but at the expense of increased variance.

Overall, based on the simulation results, serial correlation of the magnitude we might expect in stock returns does not seem to be a major concern when estimating the long-run distribution of returns. This is especially true when one considers the overall estimation uncertainty, which tends to dwarf the empirical differences between most estimation methods.

4 Analyzing long-run returns with panel data

So far, we have focused on using the return series of a single asset or market to estimate long-run returns. We now extend the analysis to a panel setting, where return series from multiple assets are available. A prominent example, also used in our empirical application, is a panel of international stock indexes.

Such data naturally raise two questions: To what extent does access to a richer panel improve the precision of inference on the long-run return distribution? And how can we empirically detect differences in long-run return distributions across countries? We address these questions in what follows.

4.1 Pooled estimation with multiple return series

Suppose that we observe K different return series and let $y_{i,t}$ denote the log return on asset i at time t. For clarity, suppose henceforth that $i = 1, ..., K$ is a country index and that the time index $t = 1, ..., n$ corresponds to different months. Suppose that all return series are i.i.d. across time and share the same marginal distribution. We do not impose any further structure on the joint distribution of the vector of cross-sectional returns $\mathbf{y}_t = (y_{1,t}, \ldots, y_{K,t})^\top$. In particular, we allow for cross-sectional correlation due to shared factors.

Estimators of the shared mean μ_y, variance σ_y^2, and skewness γ_y of the one-period log

returns are given by the pooled estimators,

$$\hat{\mu}_y^{\text{pool}} = \frac{1}{nK}\sum_{i=1}^{K}\sum_{t=1}^{n} y_{i,t}, \quad \hat{\sigma}_y^{2,\text{pool}} = \frac{1}{nK}\sum_{i=1}^{K}\sum_{t=1}^{n}\left(y_{i,t} - \hat{\mu}_y^{\text{pool}}\right)^2, \tag{30}$$

and

$$\hat{\gamma}_y^{\text{pool}} = \frac{1}{\hat{\sigma}_y^{3,\text{pool}} nK}\sum_{i=1}^{K}\sum_{t=1}^{n}\left(y_{i,t} - \hat{\mu}_y^{\text{pool}}\right)^3. \tag{31}$$

Based on these estimators, we obtain a pooled skewness-corrected ML estimator

$$\widehat{Q}_{Y_T}^{\text{ML-skew,pool}}(\tau) = T\hat{\mu}_y^{\text{pool}} + \sqrt{T}\hat{\sigma}_y^{\text{pool}}\Phi^{-1}(\tau) + \frac{1}{6}\hat{\sigma}_y^{\text{pool}}\hat{\gamma}_y^{\text{pool}}\left(\left(\Phi^{-1}(\tau)\right)^2 - 1\right). \tag{32}$$

A bespoke estimator for the log-normal model $\widehat{Q}_{Y_T}^{\text{ML,pool}}$ is obtained as a special case of this formula by setting $\hat{\gamma}_y^{\text{pool}}$ equal to zero.[12]

A panel-version of the FF bootstrap can be implemented by simply resampling from the entire pooled data set; we refer to this estimator as $\widehat{Q}_{Y_T}^{\text{boot,pool}}(\tau)$.[13] For simplicity, in the panel setting we restrict the formal analysis to the skewness-corrected MLE. Both the simulations (reported in the Online Appendix) and the empirical analysis show that the pooled skewness-corrected ML and the pooled FF bootstrap estimators yield very similar results.

To construct a confidence interval around the pooled estimator $\widehat{Q}_{Y_T}^{\text{ML-skew,pool}}$, let

$$\widehat{V}_t(\tau) = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{y_{i,t}}{\hat{\sigma}_y^{\text{pool}}} + \frac{1}{2\sqrt{T}}\left(\frac{y_{i,t} - \hat{\mu}_y^{\text{pool}}}{\hat{\sigma}_y^{\text{pool}}}\right)^2\Phi^{-1}(\tau)\right), \tag{33}$$

and

$$\hat{\sigma}_{V,\tau}^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\widehat{V}_t(\tau) - \frac{1}{n}\sum_{s=1}^{n}\widehat{V}_s(\tau)\right)^2. \tag{34}$$

[12]We label the pooled estimator for the normal case with an "ML" moniker and refer to it as the pooled or panel MLE to emphasize that it is a counterpart of the time-series ML estimator. In the presence of cross-sectional dependence, it is not numerically identical to the true maximum likelihood estimator of the panel model.

[13]ACO pursue a block bootstrap version of this in their analysis.

For the τ-quantile of the long-run log return, the lower and upper bounds of a $(1-\alpha)$-confidence interval are given by

$$L_{Y_T}^{\text{pool}}(\tau;\alpha) = \widehat{Q}_{Y_T}^{\text{ML-skew,pool}}(\tau;\alpha) - \frac{\sqrt{n}}{T}\hat{\sigma}_{V,\tau}\hat{\sigma}_y^{\text{pool}}\Phi^{-1}(1-\alpha/2), \tag{35}$$

$$U_{Y_T}^{\text{pool}}(\tau;\alpha) = \widehat{Q}_{Y_T}^{\text{ML-skew,pool}}(\tau;\alpha) + \frac{\sqrt{n}}{T}\hat{\sigma}_{V,\tau}\hat{\sigma}_y^{\text{pool}}\Phi^{-1}(1-\alpha/2). \tag{36}$$

We provide an asymptotic justification of this interval in the Online Appendix. A corresponding confidence interval for the quantiles of the gross long-run returns can be obtained by exponentiating the bounds for the log return. As in the time-series case, the confidence interval can be recentered around the ML estimator to obtain an asymptotically valid interval when the single-period log returns are non-skewed (i.e., when $\gamma_y = 0$).

4.2 The precision gains from pooling

To compare the panel confidence interval to the interval based on a single time series, consider an illustrative example with a simple factor model given by

$$y_{i,t} = \mu_y + z_t + \epsilon_{i,t}, \tag{37}$$

where z_t is a zero-mean common factor affecting all countries and $\epsilon_{i,t}$ is an idiosyncratic error term. Let $\lambda = \text{var}(z_t)/\sigma_y^2$ denote the fraction of the variance in the single-period log returns that is explained by the common factor. Empirically, this fraction can be estimated using the analysis of variance formula,

$$\hat{\lambda} = 1 - \frac{\sum_{i=1}^{K}\sum_{t=1}^{n}\left(y_{t,i} - \bar{y}_t\right)^2}{\sum_{i=1}^{K}\sum_{t=1}^{n}\left(y_{t,i} - \hat{\mu}_y^{\text{pool}}\right)^2}, \tag{38}$$

where $\bar{y}_t = \frac{1}{K}\sum_{i=1}^{K} y_{t,i}$. This expression suggests the interpretation of λ as the R^2 in a CAPM regression. For international index data, the world CAPM has decent explanatory power and an R^2 of at least 30% is often observed (see, for instance, Ferson and Harvey, 1994). In our empirical analysis, we estimate $\lambda = 0.396$ in a global panel of returns.

In the Online Appendix, we show that, up to higher-order terms,

$$\text{var}\left(\widehat{V}_t(\tau)\right) \approx \text{var}\left(z_t\right)/\sigma_y^2 + \frac{\text{var}\left(\epsilon_{i,t}\right)/\sigma_y^2}{K} = \lambda + \frac{1-\lambda}{K}. \tag{39}$$

Noting that $\hat{\sigma}_y \approx \sigma_y \approx \hat{\sigma}_y^{\text{pool}}$ and $\psi_T(\tau) \approx 1$ if T is large, the length of the confidence interval based on the panel data (see equations (35) and (36)) is therefore approximately given by

$$\begin{aligned} U_{Y_T}^{\text{pool}}(\tau;\alpha) - L_{Y_T}^{\text{pool}}(\tau;\alpha) \approx & \left(U_{Y_T}(\tau;\alpha) - L_{Y_T}(\tau;\alpha)\right)\sqrt{\lambda + (1-\lambda)/K} \\ = & \frac{2T}{\sqrt{n^*}}\hat{\sigma}_y \Phi^{-1}(1-\alpha/2)\psi_T(\tau), \end{aligned} \tag{40}$$

where

$$n^* = \left(\frac{K}{\lambda(K-1)+1}\right) n. \tag{41}$$

Thus, the length of the panel confidence interval is roughly as long as a confidence interval based on a single time series of length n^* (see equations (12) and (13)). We can interpret n^* as a measure of the *effective sample size* in the panel, accounting for the information loss due to the common factor. We require a time-series observed over n^* time periods to obtain the same precision as a K-country panel observed over n time periods. For example, if $\lambda = 0.4$ and we observe $K = 20$ countries over 120 years of monthly data ($n = 1440$), then the panel is about as precise as a single time series of $n^* \approx 2.33 \times 1,440 \approx 3,350$ observations. Compared to using a single time-series, the effective sample size increases only by a factor of 2.33, even though the number of observations used for estimation has increased twenty fold.

The effective sample size increases in K, but is bounded above by n/λ. That is, for $\lambda = 0.3$, n^* is at most $n^* \approx 3.3 \times n$, and for $\lambda = 0.4$, n^* is at most $n^* = 2.5 \times n$. In many empirical settings, λ is likely to exceed 0.3. Under such circumstances, the panel gains are limited to an equivalent of about three independent time series, regardless of the size of the cross section.

4.3 Testing for distributional differences

We now discuss a formal test for differences in the long-run distribution between one focal country $i = 1$ and a pool of homogeneous countries $i = 2, \ldots, K$. In particular, we assume that we observe i.i.d. (across time) observations $\mathbf{y}_t = (y_{t,1}, y_{t,2}, \ldots, y_{t,K})$ where the marginal log returns $y_{t,2}, \ldots, y_{t,K}$ share the same distribution $Q_{Y_T,-1}$ with common mean $\mu_{y,-1}$, variance $\sigma^2_{y,-1}$, and skewness $\gamma_{y,-1}$. The marginal distribution of country 1 is $Q_{Y_T,1}$ with mean $\mu_{y,1}$, variance $\sigma^2_{y,1}$, and skewness $\gamma_{y,1}$. As above, we do not restrict the cross-sectional correlations of the components of $\mathbf{y}_t$. In particular, the focal country and the pool of comparison countries may be exposed to the same factors. The test also applies in the special case when $K = 2$, such that the distribution of one country is tested against the distribution of another single country.

We test for differences in the τ-quantiles of the long-run return distributions, $Q_{Y_T,1}(\tau)$ and $Q_{Y_T,-1}(\tau)$. In particular, we test the null hypothesis

$$H_0 : Q_{Y_T,1}(\tau) - Q_{Y_T,-1}(\tau) = \Delta(\tau), \tag{42}$$

where $\Delta(\tau)$ is a prespecified difference between the focal country and the pool of comparison countries. Setting $\Delta(\tau) = 0$ tests for equality of the τ-quantiles. The alternative hypothesis is simply $H_1 : Q_{Y_T,1}(\tau) - Q_{Y_T,-1}(\tau) \neq \Delta(\tau)$.

To obtain an estimate of the τ-quantile for country 1, we let $\widehat{Q}^{\text{ML-skew}}_{Y_T,1}(\tau)$ denote the skewness-corrected MLE based on the return series of country 1. To estimate the τ-quantile for the other countries, we use the correspodning leave-one-out pooled skewness-corrected MLE that excludes country one. We denote this estimator by $\widehat{Q}^{\text{ML-skew,pool}}_{Y_T,-1}(\tau)$.[14] Our test statistic is based on the estimated difference between country one and the other countries,

$$T_\Delta(\tau) = \frac{\sqrt{n}}{T\hat{\sigma}_{W,\tau}} \left(\widehat{Q}^{\text{ML-skew}}_{Y_T,1}(\tau) - \widehat{Q}^{\text{ML-skew,pool}}_{Y_T,-1}(\tau) - \Delta(\tau) \right), \tag{43}$$

where $\hat{\sigma}_{W,\tau}$ is an estimator of the variance of the estimated difference that is specified in the

[14] In the $K = 2$ case, where two individual countries are tested against each other, this pooled estimator simplifies to the time-series estimator for country 2.

Online Appendix. The null hypothesis is rejected if the absolute value of the test statistic exceeds a critical value that is computed from the standard normal distribution; that is, reject at the 5%-level if $T_\Delta(\tau) \geq 1.96$ or $T_\Delta(\tau) \leq -1.96$. Additional details, including a formal asymptotic justification of this test, are found in the Online Appendix.

5 Simulation results

In this section, we present simulation results. Monthly period returns are generated according to a number of different models and we evaluate the performance of the empirical estimators described in previous sections. We focus on the case where a single time series of returns is available for inference. This setup captures the situation where one is attempting to determine the empirical distribution for a given country's aggregate stock market returns. In the Online Appendix, we present simulation results for a panel data setting, where multiple time-series are available. The panel simulation results confirm the theoretical predictions in Section 4 and the results are qualitatively very similar to those presented for the time-series case.

Additional simulations in the Online Appendix show results for estimates based on a smaller sample size (60 years of data rather than 120 years), which naturally leads to less precise estimates. Moreover, the effect of changing the sampling frequency is evaluated, and results show that the precision of the quantile estimates is virtually identical for daily, monthly, and annual data; there are thus no gains to using higher-frequency data.

5.1 Return distributions

We simulate four different data generating processes (dgps), two with i.i.d. returns (log-normal, log-normal-with-crashes) and two with non-i.i.d. returns (stochastic volatility, long-term mean reversion).

- *Log-normal:* Gross period returns x_t are i.i.d. log-normal and log returns y_t are thus i.i.d. normal. This specification is completely determined by the mean, μ_y, and the variance, σ_y^2, of the log returns. In our notation, we parameterize both the

log-normal and normal distributions with μ_y and σ_y^2, such that, $x_t \sim LN\left(\mu_y, \sigma_y^2\right)$ and $y_t \sim N\left(\mu_y, \sigma_y^2\right)$.[15]

- *Log-normal-with-crashes:* The distribution of y_t is given by a mixture of two normals that are i.i.d. over time, and the gross-returns are given by a mixture of two log-normals,

$$x_t \sim \begin{cases} LN\left(\tilde{\mu}_y, \tilde{\sigma}_y^2\right) \text{ with probability } 1-p \\ LN\left(\kappa + \tilde{\mu}_y, \tilde{\sigma}_y^2\right) \text{ with probability } p \end{cases}. \tag{44}$$

  If $\kappa < 0$, and $|\kappa| >> \tilde{\sigma}_y$, the p-probability outcome corresponds to a (low-probability) crash.

- *Stochastic volatility:* The gross-returns x_t are generated by discretely sampling a continuous time stochastic volatility (SV) process. In particular, x_t is generated from a continuous time diffusion model with a separate Brownian motion driving the stochastic volatility process. The SV model is a deviation from the i.i.d. assumption of the theoretical analysis and the period returns generated from it follow a martingale difference sequence (m.d.s.) rather than an i.i.d. sequence.

- *Long-term reversals:* Log returns y_t follow a moving average process of order q (MA(q)) with negative MA coefficients and i.i.d. normal innovations. This leads to returns that exhibit a q-month reversal effect. The gross return process is the exponentially transformed linear MA process.

All distributions are parameterized to have identical (unconditional) mean μ and variance σ^2.[16] The two moments are calibrated to aggregate market returns with the average monthly returns μ set to 0.6% ($\mu = 1.006$) and the volatility set to 6% ($\sigma = 0.06$). These values are similar to those reported for the global real returns used in the subsequent empirical analysis in Section 6 (and also similar to those reported in ACO; they report

[15] Let μ and σ^2 denote the mean and variance of the *gross* returns. When gross returns are log-normal (and log returns normal) the mean and variance of the *log* returns are then given by $\mu_y = \log\left(\frac{\mu^2}{\sqrt{\sigma^2+\mu^2}}\right)$ and $\sigma_y^2 = \log\left(\frac{\sigma^2}{\mu^2}+1\right)$.

[16] The MA-specification has the same mean and variance as the other processes, but its *long-run* variance (and as a consequence the long-run mean) differs from the other specifications.

that monthly real returns in their global sample have a mean of 0.55% and a volatility of 5.86%).

The parameters of the log-normal distribution are completely determined by μ and σ. In the log-normal-with-crashes specification, we set $p = 1/100$ and $e^{\kappa} = 0.7$, which implies that crashes occur with 1% probability each month (on average one crash every 8 years) and that each crash corresponds to a 30% expected loss; a complete specification is given in the Online Appendix.

The SV model follows a fairly standard parameterization, apart from the fact that we set the mean gross return to μ and the average volatility to σ. The correlation between the price process and the stochastic volatility process is set to -0.5, which represents a sizable but empirically reasonable so-called leverage effect.[17]

In the long-term reversal process, we set $q = 60$ (i.e., log returns follow an MA(60) process), to capture a 5-year reversal effect in returns. Whether such reversals truly exists remain a topic of some controversy. Here we simply simulate returns consistent with such effects and evaluate the impact on the estimates of long-run return distributions.[18] We parameterize the MA process such that the returns have a variance ratio of 0.8. That is, in the long run returns have 20 percent less variance than in the short run, which represents a substantial deviation from the baseline random walk assumption; again, a complete specification of the implementation is provided in the Online Appendix.[19]

We focus on compounding horizons of 10 and 30 years ($T = 120$ and $T = 360$).

5.2 Point estimate precision and confidence interval coverage

For each data-generating process (e.g., log-normal), we generate 10,000 sample paths with $n = 1,440$ monthly returns, representing 120 years of data. For each simulated sample path, for a given horizon T and quantile τ, the ML ($\widehat{Q}^{\mathrm{ML}}_{X_T}(\tau)$), the skewness-corrected ML

[17] Our simulation implementation for the stochastic volatility distribution follows Farago and Hjalmarsson (2023a), who provide more details.

[18] There is a large literature evaluating long-term mean reversion in stock returns. The findings in Fama and French (1988), Poterba and Summers (1988), Cecchetti et al. (1990), Cutler et al., (1991), Siegel (2008), and Spierdijk et al. (2012), generally support long-term mean reversion. Several other studies, including Richardson and Stock (1989), Kim et al. (1991), and Richardson (1993), are more negative.

[19] In brief, the MA coefficients, θ_k, $k = 1, ..., q = 60$, are assumed to decline for greater lags. We use the parametric form $\theta_k = \frac{\theta_1}{\sqrt{k}}$, where the value of θ_1 is set to achieve a variance ratio of 0.8.

($\widehat{Q}_{X_T}^{\text{ML-skew}}(\tau)$), and the FF bootstrap ($\widehat{Q}_{X_T}^{\text{boot}}(\tau)$) estimates of quantile τ of the long-run distribution of returns are calculated, along with corresponding confidence intervals.

The results are shown in Figures 1 and 2 and Tables 1 and 2, for $T = 120$ and $T = 360$, respectively. The figures and tables follow the same layout, with Panels A and B showing results for the (i.i.d.) log-normal and log-normal-with-crashes models, respectively, and Panels C and D showing results for the (non-i.i.d.) SV and long-term reversal specifications, respectively.

The figures show simulation results for all quantiles of the long-run distribution. For a simple explanation of the figures, fix a value of τ. For example, suppose that we want to estimate the 20^{th} percentile of the long-run distribution of returns ($\tau = 0.2$). Find the value 20 on the vertical axis and then move horizontally to the solid line to read the true (population) quantile off of the horizontal axis.[20] The estimators of this specific population parameter have a sampling distribution that we characterize by its simulated 5^{th}, 50^{th} (median) and 95^{th} percentiles. These percentiles are computed as the empirical percentiles of the 10,000 simulated estimates. It is important not to confuse percentile values on the vertical axis (which define the parameter of interest) with the concept of a percentile that characterizes the sampling distribution (for a given parameter of interest). For the skewness-corrected ML, the 5^{th}, 50^{th} and 95^{th} percentiles of the sampling distribution are shown as dashed lines. For the bootstrap estimator, the percentiles are shown as dashed-and-dotted lines. For the ML estimator, the median is shown as a dotted line and the 5^{th} and 95^{th} percentiles are shown by the left and right edges of the shaded region, respectively. The sampling distributions for different estimators are very similar, making the lines corresponding to different estimators difficult to distinguish in the figures.

Starting from the value 20 on the vertical axis and then moving left to right, we first encounter the 5^{th} percentile, then the median and then the 95^{th} percentile. For example, in Panel A of Figure 1, the true value of the 20^{th} percentile of the 10-year return is 0.96. The estimates fall between 0.7 (5^{th} percentile of the sampling distribution) and 1.3 (95^{th} percentile of the sampling distribution), 90 percent of the time. The *horizontal* distance

[20] We use simulations to obtain very precise approximations of the true quantiles.

between the 5^{th} and 95^{th} percentiles gives an indication of the precision of the estimator.

The tables provide additional information for a selection of quantiles. In particular, the tables show the median error (i.e., median bias) and the median absolute error of each estimator. As a result of the exponential transformation (from log to gross returns), the estimates of the quantiles of long-run compound gross returns tend to be mean biased but median unbiased in the benchmark log-normal case. Given that the bias is primarily induced by the non-linear exponential transformation, the median errors are therefore more informative. In the top rows of each panel in the tables, the population quantiles are shown together with the actual average coverage rates of nominal 90% confidence intervals for each estimator.

We first discuss simulation results for the i.i.d. specifications, starting with the benchmark log-normal model (Panel A of each figure and table). The sampling distributions for all three estimators are virtually identical, both for $T = 120$ and $T = 360$. This is evidenced by the percentiles of the sampling distributions plotted in Figures 1 and 2 and the median and median absolute errors in Tables 1 and 2, which are essentially identical for all estimators across all quantiles. In particular, there is no loss of precision from using the more generally valid skewness-corrected ML or bootstrap estimators, as opposed to the (correctly specified) MLE. The shared sampling distribution is median-unbiased for the true population quantiles, such that our estimators recover the true long-run return distribution "on average".

The empirical coverage rate of the confidence intervals, as simulated by the average coverage rate over the 10,000 simulations, is close to the nominal 90% level for all quantiles τ and all three approaches. The confidence intervals thus reliably capture the sampling uncertainty of the estimators.

This sampling uncertainty is substantial and, as predicted by our theory, even larger for the longer horizon of $T = 360$ (30 years) than for $T = 120$ (10 years). For instance, for the 10^{th} percentile of the long-run distribution, the median absolute error increases from 0.09 for the 10-year horizon to 0.39 for the 30-year horizon (see Panel A in Tables 1 and 2). For the 30-year horizon, the margin of sampling uncertainty dwarfs the magnitude of the

true values. For example, the true value of the 10^{th} percentile of the 30-year distribution is equal to 1.07 (= 7% net gain), but the range of 90% probability outcomes in Panel A in Figure 2 stretches from 0.4 (= 60% net loss) to 2.7 (= 170% net gain). Clearly, interpreting the raw point estimates without considering sampling uncertainty can be highly misleading.

We next turn to simulation results for the log-normal-with-crashes model (Panel B of each figure and table). This return distribution is heavier-tailed and exhibits substantial negative skewness that is driven by the rare crashes. Here, the ML estimator is misspecified and exhibits a small bias when estimating the upper percentiles of the long-run return distribution (as seen in Tables 1 and 2). As expected from the theory, the skewness-corrected ML and bootstrap estimators are mostly median-unbiased and have similar sampling distributions. All confidence intervals for the 10^{th} percentile or higher have empirical coverage rates close to the nominal 90% level ($\pm$4%). For the lowest percentiles, the coverage is somewhat poor at the shorter 10-year horizon: the coverage rates for the 1^{st} percentile of the 10-year return are 78% for the MLE and 80% for the skewness-corrected MLE and bootstrap estimator. In line with our large-T justification of the confidence intervals, the coverage improves for the 30-year horizon, where the coverage rates are always above 85%.

Turning to the non-i.i.d. distributions, we first consider the stochastic volatility (SV) model (Panel C of each table and figure). This specification is one of the workhorse models for modeling short-term stock return data and captures stylized features, such as time-varying volatility and leverage. However, when returns are compounded over longer horizons, these features do not appear to have a major impact. The MLE is able to capture the long-run distribution quite well (in a median-unbiased sense). The bias-correction provided by the skewness-corrected ML and bootstrap estimators is only noticeable at the 95^{th} percentile (or higher) of the long-run distribution. Apart from this, the three estimators behave very similarly and have virtually identical median absolute errors. The empirical coverage rates for the confidence intervals are close to the nominal 90% level, with the exception of the lowest percentiles when considering 10-year returns. Similarly

to the log-normal-with-crashes model, the coverage rates for the lowest percentiles are considerably improved for the longer 30-year horizon (where they are always above 85%).

The long-term reversals specification introduces serial correlation (Panel D of each table and figure). The non-i.i.d. aspect of this specification does not get washed out in the long run: by construction the process has a long-run variance that is different from its short-run variance (the ratio between the two variances is set to 0.8). Even though none of our three estimation approaches control for serial dependence, they still perform well. The bias is slightly larger than for the other specifications, but the coverage rates for the 90% confidence intervals are still good ($\geq 85\%$). Reliable inference in the presence of serial correlation is therefore still possible with the estimators derived here.

6 Empirical analysis of global stock returns

6.1 The DMS data set

We now turn to an an empirical analysis of the long-run distribution of aggregate stock market returns. Our data source is the panel of annual international stock returns described in Dimson, Marsh, and Staunton (2021), and subsequently referred to as the DMS data set. This data set is an updated and extended version of the panel of 21 countries constructed and explained in detail by Dimson, Marsh, and Staunton (2002).

The data set contains uninterrupted return series from 1900 to 2020 for the 21 countries originally included in the 2002 DMS data; see Table 3 for a list of countries. In addition, there are returns for 11 countries for which the data start later than 1900. In our main analysis, we focus on the 21 countries with a full history, since an identical sample period allows for the cleanest cross-country and global-to-country comparisons. Below, when we refer to the full panel of global returns, we refer to this 21-country panel. In the Online Appendix, we present empirical results based on all 32 countries; the results from this extended data set are very similar to those shown for our main specification, and do not change any of our qualitative conclusions.[21]

[21]Figure A8 in the Online Appendix provides an overview of data availability for each country in the extended data set.

The return data are reported at an annual frequency. Our sample size is therefore measured in years ($n = 121$ for the countries will a full history from 1900), and as before we consider compounding horizons spanning $T = 10$ and $T = 30$ years. As discussed in Section 3.3, and verified by simulations in the Online Appendix, whether one samples returns on an annual or a monthly frequency has virtually no impact on the precision of the estimates of the long-run distribution. Rather, it is the total time span of the data (i.e., the number of years covered), that is the key determinant of estimation precision.

Throughout the analysis, we focus on real returns expressed in local currency, which are provided directly in the DMS database. Table 3 shows descriptive statistics for the one-period annual returns for each of the 21 countries in the sample, as well as for the full panel pooling the return series from all these countries. The arithmetic (simple) mean returns range from 5.02% to 9.31% per year, while the geometric mean ranges from 0.86% to 7.06%. The relatively large differences between the arithmetic and geometric means are mostly attributable to the presence of large negative returns, as evidenced by the minimum values also presented in the table. The standard deviations of the return series are between 16.73% and 33.64%, which is typical for yearly returns of equity indexes. The full panel has an average return of 7.32% (4.63%) using the arithmetic (geometric) mean and a volatility of 24.03%.

6.2 Long-run distributions of global real returns

We first consider the global long-run return distribution. To estimate the distribution of global long-run returns, we pool the data from the panel of 21 countries with a full history in the DMS data set. The three pooled estimators described in Section 4 are implemented; that is, the normal ML ($\widehat{Q}_{X_T}^{\text{ML,pool}}(\tau)$), the skewness-corrected ML ($\widehat{Q}_{X_T}^{\text{ML-skew,pool}}(\tau)$), and the FF bootstrap ($\widehat{Q}_{X_T}^{\text{boot,pool}}(\tau)$) estimators. Our implementation of the FF bootstrap estimator follows the description in Section 3.4.2, with the number of bootstrap repetitions set to $B = 100,000$, sampling from the the pooled returns data from the 21 countries in the panel. To quantify sampling uncertainty, confidence intervals are calculated from equations (35) and (36). The confidence intervals are centered on each of the three pooled

estimators.[22,23]

The informativeness of the pooled data depends on the extent to which the return series are correlated. Our confidence intervals account for cross-sectional dependence in a nonparametric way, without imposing any specific structure such as, for example, a factor model. To provide an intuitive summary measure of the informativeness of the panel, the effective sample size n^* in equation (41) is also calculated. n^* is derived under the more restrictive assumption of a one-factor model, but provides an easily interpreted measure that is also empirically relevant.[24]

To compute the effective sample size, we estimate λ based on equation (38), using the full sample of 21 countries spanning the period from 1900 to 2020. This gives an estimate $\hat{\lambda}_{\text{Full sample}} = 0.396$, indicating that approximately 40% of the variation in returns can be explained by a common factor in the cross section. Based on this estimate for λ, the effective number of time-series observations for the panel with $K = 21$ countries and $n = 121$ years is

$$n^*_{\text{Full sample}} = \frac{21}{0.396 \times 20 + 1} \times 121 = 2.354 \times 121 \approx 285.$$

In other words, the 21-country panel, containing a total of $121 \times 21 = 2541$ annual observations, is equivalent in estimation precision to a single time series with 285 years of data.

Figure 3 and Table 4 show the empirical results for the long-run global real returns. Figure 3 follows a similar format to Figures 1 and 2, which show results for the simulation

[22]As pointed out in Section 4.1, in the panel case we only formally validate the confidence intervals around the pooled (skewness corrected) ML estimator, not around the pooled FF bootstrap estimator. As seen from the empirical results, as well as the simulation results in the Online Appendix, the pooled FF bootstrap estimator is very similar to the pooled skewness-corrected ML estimator.

[23]The theoretical results provide analytical confidence intervals for the log returns, which are then simply exponentiated to achieve confidence intervals for the gross returns. The length of the confidence intervals for the *log* returns is identical irrespective of the estimation method used for calculating the point estimates. The length of the confidence intervals around the *gross* returns might differ slightly across the different estimation methods, but these differences are solely due to the non-linearity of the exponential transformation.

[24]In non-reported results, we find that confidence intervals, calculated by treating the pooled data as a single time series with sample size n^*, yield a good approximation of our more generally valid confidence intervals. This is not terrible surprising since the one-factor world CAPM, while far from a perfect model, still has significant explanatory power for global index returns (e.g., Ferson and Harvey, 1994).

exercises. Table 4 tabulates, for selected quantiles, the results shown in Figure 3. In line with the theoretical analysis, as well as the simulation results, the tables and figures present empirical estimates of long-run *gross* returns. When discussing the results in the text, we translate the *gross* returns into *net* returns, expressed in percent, for ease of interpretation. For instance, in Panel A of Table 4, the full-sample ML estimate of the median of the 10-year gross return is found to be 1.57. In the text, we report this as a net return of 57%.

Panels A1 and B1 in Figure 3 present empirical estimates for compounding horizons of 10 and 30 years, respectively, using the full sample period from 1900 to 2020. The dotted line and shaded area in each panel are the point estimate and the 90 percent confidence band, respectively, for the different quantiles based on the pooled ML estimator. The dashed and the dash-dotted lines represent the pooled skewness-corrected ML and FF bootstrap estimates, respectively, along with the corresponding 90 percent confidence bands. There are some slight differences in the point estimates for the three methods for the $T = 10$-year compounding horizon (Panel A1 in Figure 3). However, the differences across the point estimates are small compared to the uncertainty of each estimate, represented by their confidence bands. For example, the estimates of the median 10-year net return are 57%, 64% and 62% according to the MLE, the skewness-corrected MLE and the FF bootstrap estimator, respectively. The confidence interval for the median 10-year return (based on the skewness-corrected MLE) stretches from 31% to 107%; exact numerical values are found in Table 4.

The differences between the estimators are even less noticeable for the longer 30-year compounding horizon (Panel B1 in Figure 3). The confidence interval for the median net return, based on the skewness-corrected MLE, now covers the range from 104% to 709% and the differences between the three point estimates are arguably negligible relative to this large uncertainty in the estimates. In general, the uncertainty around the 30-year returns is much larger than that around the 10-year returns, indicating the difficulty to empirically pin down the distribution of very long-run returns.

Panels A2 and B2 in Figure 3 show estimates using the same panel of countries, but

restricting the sample to the years from 1960 to 2020. To the extent that one is unwilling to draw strong inference from data in the distant past, this more recent sample is of great interest (FF restrict their empirical analysis for the U.S. to start in 1963). The post-1960 sample roughly corresponds to an era of more modern, as well as globally more integrated, financial markets. This greater integration is reflected in the correlation of returns across countries. The estimated value for λ, measuring cross-sectional correlation, is $\hat{\lambda}_{\text{Post-1960}} = 0.527$ for the post-1960 sample, which yields an effective sample size of $n^*_{\text{Post-1960}} = 111$. That is, the full panel between 1960 and 2020, containing $61 \times 21 = 1281$ annual observations, barely dominates a single time series of 100 years of returns in terms of estimation precision.

As a direct consequence of the shorter time series, and hence a much smaller sample size, the estimation uncertainty substantially increases. This is particularly noticeable for the longer 30-year horizon, as seen from comparing the full sample results and the post-1960 results in Panels B1 and B2 of Figure 3. In the post-1960 sample, the median net 30-year return can range from 66% to 1484%, whereas the corresponding range in the full sample is from 104% to 709%, based on the 90 percent confidence intervals around the skewness-corrected MLEs.[25]

6.3 Individual country returns

We next look at the long-run distribution for each country separately. Since the different estimation methods lead to very similar outcomes, we only present results for the skewness-corrected MLE. Estimation results for all three methods are tabulated in Tables A4 to A9 in the Online Appendix.

In the discussion of the country-specific results, our main focus is on comparisons between the (pooled) global and individual country estimates. We try to assess whether country-specific long-run returns are best viewed as the outcomes of unique processes, or if they could be viewed as realizations from the same (global) data generating process.

[25]Table A10 in the Online Appendix repeats the results presented in Table 4, but using all 32 countries in the extended DMS data set. The quantitative results are very similar and the overall conclusions do not change when using the larger unbalanced panel.

6.3.1 U.S. versus global

We start our analysis with considering the long-run stock returns for the U.S.. These are of particular interest, both because the U.S. has by far the largest market capitalization among all countries, and because of the historically strong performance of the U.S. market. The latter phenomenon is sometimes posited as the U.S. being a "lucky survivor" and that its performance is unlikely to be representative of other countries (Goetzmann and Jorion, 1999, and van Binsbergen et al., 2023). Figure 4 contains a visual comparison of global and U.S. long-run returns. Using the full sample from 1900 to 2020, Panel A1 shows results for the 10-year return distribution and Panel B1 shows results for the 30-year distribution. The global estimates are formed using the pooled panel sample (including the U.S.), whereas the U.S. estimates are based solely on the time-series of returns for the U.S.. The dashed lines represent the global point estimates, with the shaded area capturing the 90 percent confidence band. The solid lines are the U.S. point estimates, along with 90 percent confidence bands represented by the dotted lines. In addition, p-values from the test in Section 4.3—for equality of the global and the U.S. percentiles—are displayed in the lower-right corners of the graphs, for the 5^{th}, 50^{th}, and 95^{th} percentiles.

An immediate observation is that, for both return horizons (Panels A1 and B1), the point estimates of the U.S. distributions of long-run returns are to the right of the global distributions, for all the percentiles shown in the graphs. The point estimate for the 5^{th} percentile of the 30-year returns is a 55% net loss globally, but a 14% gain in the U.S.. This is a numerically large difference, but the 90 percent confidence interval for the 5^{th} percentile of the global net returns stretches from -80% to -3%, while the corresponding confidence interval for the U.S. returns stretches from -57% to 204%. There is thus great estimation uncertainty and a substantial overlap between the confidence intervals. However, the formal hypothesis test for percentile equality shows that the 5^{th} percentile for the 30-year U.S. and global returns are statistically significantly different with a p-value of 0.03.[26]

[26]There is a seeming discrepancy between the great uncertainty in the estimates for the U.S. percentiles and the ability of the formal test to reject the hypothesis of equality. This tension is explored in the Online Appendix, where we illustrate how estimates of the *differences* between percentiles can be more precise than the estimates of the actual percentiles, resulting in formal hypothesis tests that have greater power than what one might expect based on the confidence intervals around each estimate.

For the median net returns at the 30-year horizon, the global estimate is 307% and the U.S. estimate is 597%. At face value, these appear very different, but again the estimation uncertainty is great. The U.S. point estimate firmly lies within the 90% confidence interval for the global median, which stretches from 104% to 709%. The p-value for the test of equality is equal to 0.17 and we cannot reject that the global and U.S. medians are the same. (All numerical values can be found in Tables A6 and A7 in the Online Appendix.)

The confidence bounds are uniformly narrower for the shorter 10-year horizon, compared to the 30-year horizon. Despite the higher precision for the shorter horizon, the confidence intervals for the global and U.S. returns substantially overlap for all percentiles. The conclusions from the formal tests are identical for the 10-year horizon: the null hypothesis of equality for the U.S. and the global percentiles are rejected for the 5^{th} percentile (p-value of 0.01), but not for the median or the 95^{th} percentile.

We also note that the confidence bands for the U.S. are wider than for global returns, which is simply a result of less data: when calculating the U.S. returns, we only have one time series at our disposal, in contrast to the global returns, where we can make use of a full panel of returns.

Panels A2 and B2 of Figure 4 present results based on the post-1960 sample. The qualitative conclusions from this shorter sample are essentially the same as those we draw from the full sample. There is now considerably lower precision in the global pooled estimates (recall that the effective sample size, n^*, for this panel is equal to 111, and thus offers less precision than the full 121-year U.S. sample used in Panels A1 and B1), resulting in very considerable overlaps between the global and U.S. confidence intervals. Overall, in this shorter sample, the U.S. distributions appear even more similar to the global ones, especially for the longer 30-year horizon. For lower percentiles, the U.S. distributions are still to the right of the global distributions, but the distributions now cross at lower percentiles (around the 65^{th} and 75^{th} percentiles, for the 10- and 30-year horizons, respectively). One can still reject the null hypothesis that the 5^{th} percentiles are identical for the 10-year U.S. and global distributions (p-value of 0.01), although for the 30-year horizon the p-value is now 0.07, such that the result is only borderline

significant. As in the full sample case, there is no evidence of statistically significantly different medians.[27]

Viewed in isolation, the results in Figure 4 offer some support for the notion that the U.S. stock market might not be representative for the global long-horizon return distribution. The evidence is concentrated to the lowest percentiles and distinguishing the central parts of the distributions (e.g., the medians) is not possible. However, by focusing on the U.S. versus the global experience, one is singling out a country that *ex post* had a strong performance, and then testing whether that country is different. In the following analysis, we therefore perform analogous tests for all countries in our data set, and calculate test statistics that also take into account the resulting multiple testing issue.

6.3.2 Other countries versus global

Figure 5 offers a concise comparison of the results for the 21 countries and the global panel, using the full sample period from 1900 to 2020.[28]

The horizontal lines in Panel A1 (B1) in Figure 5 show 90% confidence intervals for the 5^{th} percentile of the 10-year (30-year) returns for each country, with the point estimates indicated by solid circles. The estimates are all based on the skewness-corrected MLE. The corresponding global results (based on the pooled skewness-corrected MLE) for the 5^{th} percentile are shown with a vertical line (the point estimate) and a shaded area (the 90% confidence band). The p-values from the test of equality between the global percentile and each country-specific percentile are shown next to the horizontal bar for each country.[29] Although there are differences in the point estimates across different countries, there are no countries where the global and country-specific confidence intervals are disjoint. However, the formal hypothesis tests show statistically significant differences (at the 5% level) for

[27]For the 10-year horizon, there is some evidence (p-value of 0.07) that the 95^{th} percentiles are significantly different. Since the U.S. distribution is to the left of the global distribution for high percentiles, this would signal that the global 95^{th} percentile is *greater* than the U.S. one.

[28]Analogous results to the more detailed ones presented for the U.S. in Figure 4, are found for other countries in Figures A9 and A10 in the Online Appendix.

[29]The test of equality between the individual-country percentile and the global percentile uses the pooled estimate with the given country excluded (see Section 4.3). In Figures 5 and 6, as well as in Figure 4, we show the full global estimates, such that this estimate does not change depending on which country we are testing against. In practice, the full pooled estimates and the leave-one-country-out estimates are very similar.

several countries: for $T = 10$ (Panel A1), there are 11 significant outcomes and for $T = 30$ (Panel B1), there are 7 significant outcomes. The p-values for the U.S. (equal to 0.01 and 0.03 in the $T = 10$ and $T = 30$ cases, respectively) are not the smallest ones observed.

Panels A2 and B2 follow the same format as Panels A1 and B1, presenting results for the median (50^{th} percentile) estimates. In this case, few *point estimates* lie outside the global confidence interval, and there is little evidence of any statistically significant differences (the lowest p-value is equal to 0.08, for Austria at the 10-year horizon). For the upper tail of the distribution (Panels A3 and B3 show results for the 95^{th} percentiles), the only country-specific point estimate that lies outside the global confidence interval is for Canada at the 10-year horizon (with a p-value of 0.05). Otherwise, there is no evidence of statistically significant differences.

In Figure 6, corresponding results to those in Figure 5 are shown, using the shorter post-1960 subsample. The conclusions drawn from the more recent data are similar to those from the full sample, but the evidence of statistically significant differences is weaker. For the 5^{th} percentile (panels A1 and B1), 4 countries (including the U.S.) are found significantly different (at the 5% level) for the 10-year horizon, and only one country at the 30-year horizon (Italy). For the median (Panels A2 and B2), only Italy is found significantly different. For the 95^{th} percentile (Panels A3 and B3), only Canada is found significantly different.

Figure 5 (and Figure 6) confirms that the evidence of statistically significant differences is mostly concentrated to the lower percentiles. It also shows that the U.S. does not appear that unique, when compared to other countries in a symmetrical way: when compared to the global distribution of long-run returns, the 5^{th} percentile of U.S. long-run returns appear statistically significantly different, but the same holds true for a number of other countries.

However, by testing all countries, one runs into a multiple-testing issue. The individual p-values will tend to overstate the statistical significance, since we are not controlling for the fact that we conduct many simultaneous tests. Thus, the p-values less than, say, 0.05 recorded in Figure 5 cannot be viewed as evidence of statistical evidence at the 5% level.

Instead, a multiple-testing correction must be conducted. We use Holm's (1979) step-down procedure to correct for the multiple tests.[30]

Significance at the 5% and 10% levels, according to Holm's multiple testing procedure, is indicated with a dagger (†) and double-dagger (‡), respectively, next to the p-values in Figures 5 and 6. As is evident, once one controls for multiple testing, the evidence of significant differences is greatly reduced and almost eliminated. For the 30-year horizon, no statistically significant differences can be established at the 5% level, for any country or percentile, in either the full sample or the post-1960 sample. At the less stringent 10% level, the 5$^{\text{th}}$ percentile for Australia and Canada are found signigicantly different in the full sample (Panel B1 in Figure 5). For the shorter 10-year horizon, statistically significant differences at the 5% level are found for the 5$^{\text{th}}$ percentile in the full sample (Panel A1 in Figure 5) for Australia, Canada, and Denmark. In the post-1960 sample, only Canada is found statistically significant for the 5$^{\text{th}}$ percentile (Panel A1 in Figure 6). In this shorter sample, the 95$^{\text{th}}$ percentile for Canada is also found to be significantly (smaller) than the global one.

6.3.3 Summary and interpretation

To sum up, our empirical analysis reveals large uncertainty around the distribution of long-run stock returns. If one focuses on the U.S. alone, one would conclude that, for the lower percentiles, the long-run return distribution for the U.S. likely differs from the global pooled distribution. If one instead approaches the testing problem from a more agnostic data-driven viewpoint, the U.S. does not stand out and the evidence of *any* significant country-heterogeneity in long-run stock returns is quite weak, especially at the longer 30-year horizon. That is, from a purely data-driven perspective, one cannot rule out the possibility that historical long-run returns in different countries are essentially different outcomes of the same underlying global return distribution.

[30]Holm's procedure is a form of sequential Bonferroni correction. For a given significance level α, it proceeds as follows: (i) Compute the p-values, $p_1, p_2, ..., p_K$, for each individual country. (ii) Sort the countries according to their p-values, such that $p_{(1)} \leq p_{(2)} \leq ... \leq p_{(K)}$. (iii) Let $L = \min\left\{j : p_{(j)} > \alpha/(K+1-j)\right\}$. (iv) Reject the null hypothesis for all countries where $p_j < p_{(L)}$. The procedure controls the family-wise error rate (FWER) at level α.

7 Conclusion

The distribution of compound stock returns over long horizons is of great interest to private investors, as well as politicians and policy makers that manage and regulate savings in pension funds and related vehicles such as sovereign wealth funds. We study empirical estimation of such long-run distributions, and characterize the uncertainty inherent in these estimates.

The main takeaway of our analysis is that for investment horizons greater than a few years, estimation uncertainty is very large. At long horizons such as 30 years, the point estimates are almost uninformative and one should instead focus on the confidence intervals that we propose. Point estimators are formally inconsistent at longer horizons, and one needs to be careful not to overinterpret them. Apparently large differences might simply reflect statistical uncertainty, not true economic differences.

References

Anarkulova, A., S. Cederburg, and M.S. O'Doherty, 2022. Stocks for the long run? Evidence from a broad sample of developed markets. *Journal of Financial Economics* 143, 409-433.

Bessembinder, H., 2018. Do Stocks Outperform Treasury Bills? *Journal of Financial Economics* 129, 440-457.

Cecchetti, S.G., P.S. Lam, and N.C. Mark, 1990. Mean Reversion in Equilibrium Asset Prices, *American Economic Review* 80, 398-418.

Cutler, D.M., J.M. Poterba, and L.H. Summers, 1991. Speculative Dynamics. *Review of Economic Studies* 58, 529-546.

Dimson, E., P.R. Marsh, and M. Staunton, 2002. Triumph of the Optimists: 101 Years of Global Investment Returns. Princeton University Press.

Dimson, E., P.R. Marsh, and M. Staunton, 2021. Global Investment Returns Sourcebook 2021. Credit Suisse.

Fama, E., and K. French, 1988. Permanent and Temporary Components of Stock Prices. *Journal of Political Economy* 96, 246-273.

Fama, E., and K. French, 2018. Long Horizon Returns. *The Review of Asset Pricing Studies* 8, 232–252.

Farago, A., and E. Hjalmarsson, 2023a. Long-Horizon Stock Returns Are Positively Skewed. *Review of Finance* 27, 495–538.

Farago, A., and E. Hjalmarsson, 2023b. Small Rebalanced Portfolios Often Beat the Market Over Long Horizons. *Review of Asset Pricing Studies* 13, 307–342.

Ferson, W. E., and C. R. Harvey, 1994. Sources of Risk and Expected Returns in Global Equity Markets. *Journal of Banking and Finance* 18, 775—803.

Goetzmann, W.N., and P. Jorion, 1999. Global Stock Markets in the Twentieth Century. *Journal of Finance* 54, 953–980.

Hjalmarsson, E., and T. Kiss, 2022. Long-run Predictability Tests are Even Worse than You Thought, *Journal of Applied Econometrics* 37, 1334-1355.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65-70.

Kim, M. J., C. R. Nelson, and R. Startz, 1991. Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence, *Review of Economic Studies* 58, 515–528.

Merton, R.C., 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323-361.

Poterba, J., and L.H. Summers, 1988. Mean Reversion in Stock Prices: Evidence and Implications, *Journal of Financial Economics* 22, 27-59.

Richardson, M., 1993. Temporary Components of Stock Prices: A Skeptic's View, *Journal of Business and Economics Statistics* 11, 199-207.

Richardson, M., and J. Stock, 1989. Drawing Inferences from Statistics Based on Multiyear Asset Returns, *Journal of Financial Economics* 25, 323-348.

Siegel, J.J., 2008. Stocks for the Long Run. McGraw Hill, New York.

Spierdijk, L., J.A. Bikker, and P. van den Hoek, 2012. Mean Reversion in International Stock Markets: An Empirical Analysis of the 20th Century, *Journal of International Money and Finance* 31, 228-249.

Valkanov, R., 2003. Long-horizon regressions: Theoretical results and applications, *Journal of Financial Economics* 68, 201-232.

van Binsbergen, J.H., S. Hua, and J.A. Wachter, 2023. Is The United States A Lucky Survivor: A Hierarchical Bayesian Approach. Working paper, University of Pennsylvania, The Wharton School.

Table 1: Simulation results for $T = 120$.

The table shows simulation results based on estimates with a sample size of $n = 1,440$ and a compounding horizon of $T = 120$. Monthly period returns are generated according to the four different return models described in the main text: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals (Panel D). All four specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile). The next row gives the true population values of the distribution for those percentile values (i.e., the quantiles of the distribution). The following three rows show the actual average coverage rates of 90% nominal confidence bands, centered on the (i) ML estimate, (ii) skewness-corrected ML estimate, and (iii) FF bootstrap estimate. The subsequent three rows in each panel show the median error (median bias) for each of the three estimators, for a given percentile of the distribution. The final three rows in each panel show the corresponding median absolute errors (median absolute biases). The results are based on $10,000$ simulated samples.

	Panel A: Log-normal								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.36	0.57	0.72	1.07	1.66	2.57	3.82	4.85	7.56
	Coverage rates								
ML	90%	90%	90%	90%	90%	90%	90%	90%	90%
ML-skew	90%	90%	90%	90%	90%	90%	90%	90%	90%
Bootstrap	90%	90%	90%	90%	90%	90%	90%	90%	90%
	Median error								
ML	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	-0.01	-0.01
ML-skew	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	-0.01
Bootstrap	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01	-0.01
	Median absolute error								
ML	0.05	0.07	0.09	0.13	0.21	0.32	0.48	0.61	0.96
ML-skew	0.05	0.07	0.09	0.13	0.21	0.32	0.48	0.61	0.96
Bootstrap	0.05	0.07	0.09	0.13	0.21	0.32	0.48	0.61	0.96

	Panel B: Log-normal-with-crashes								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.30	0.51	0.67	1.04	1.66	2.61	3.87	4.88	7.46
	Coverage rates								
ML	78%	84%	86%	88%	90%	92%	93%	93%	92%
ML-skew	80%	84%	85%	88%	90%	92%	93%	93%	94%
Bootstrap	80%	84%	85%	88%	90%	92%	93%	93%	94%
	Median error								
ML	0.03	0.02	0.01	-0.01	-0.03	-0.02	0.06	0.16	0.58
ML-skew	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	-0.05
Bootstrap	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Median absolute error								
ML	0.06	0.08	0.10	0.15	0.22	0.32	0.47	0.59	0.97
ML-skew	0.05	0.08	0.10	0.15	0.22	0.32	0.46	0.57	0.86
Bootstrap	0.05	0.08	0.10	0.15	0.22	0.32	0.46	0.57	0.87

	Panel C: SV								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.30	0.52	0.69	1.07	1.70	2.64	3.84	4.78	7.12
	Coverage rates								
ML	74%	84%	87%	88%	89%	91%	92%	92%	90%
ML-skew	76%	85%	87%	88%	90%	91%	92%	92%	92%
Bootstrap	76%	85%	87%	89%	90%	91%	92%	92%	92%
	Median error								
ML	0.05	0.04	0.02	-0.01	-0.04	-0.05	0.02	0.13	0.58
ML-skew	0.05	0.03	0.02	0.00	-0.03	-0.04	0.01	0.09	0.41
Bootstrap	0.05	0.03	0.02	0.00	-0.03	-0.04	0.01	0.09	0.42
	Median absolute error								
ML	0.06	0.08	0.10	0.14	0.22	0.33	0.47	0.59	0.95
ML-skew	0.06	0.08	0.10	0.14	0.22	0.33	0.47	0.59	0.90
Bootstrap	0.06	0.08	0.10	0.14	0.22	0.33	0.47	0.58	0.91

	Panel D: Long-term reversals								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.41	0.62	0.77	1.11	1.66	2.48	3.56	4.42	6.64
	Coverage rates								
ML	85%	90%	91%	93%	93%	93%	91%	89%	85%
ML-skew	85%	90%	91%	93%	93%	93%	91%	89%	85%
Bootstrap	85%	90%	91%	93%	93%	93%	91%	89%	85%
	Median error								
ML	-0.05	-0.05	-0.05	-0.04	0.00	0.10	0.27	0.43	0.93
ML-skew	-0.05	-0.05	-0.05	-0.04	0.00	0.10	0.27	0.43	0.93
Bootstrap	-0.05	-0.05	-0.05	-0.04	0.00	0.10	0.26	0.43	0.92
	Median absolute error								
ML	0.06	0.08	0.09	0.13	0.19	0.29	0.45	0.60	1.05
ML-skew	0.06	0.08	0.09	0.13	0.19	0.29	0.45	0.60	1.05
Bootstrap	0.06	0.08	0.09	0.13	0.19	0.29	0.45	0.60	1.05

Table 2: Simulation results for $T = 360$.

The table shows simulation results based on estimates with a sample size of $n = 1,440$ and a compounding horizon of $T = 360$. Monthly period returns are generated according to the four different return models described in the main text: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals (Panel D). All four specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile). The next row gives the true population values of the distribution for those percentile values (i.e., the quantiles of the distribution). The following three rows show the actual average coverage rates of 90% nominal confidence bands, centered on the (i) ML estimate, (ii) skewness-corrected ML estimate, and (iii) FF bootstrap estimate. The subsequent three rows in each panel show the median error (median bias) for each of the three estimators, for a given percentile of the distribution. The final three rows in each panel show the corresponding median absolute errors (median absolute biases). The results are based on $10,000$ simulated samples.

Panel A: Log-normal									
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.33	0.71	1.07	2.12	4.55	9.75	19.36	29.20	63.09
	Coverage rates								
ML	90%	90%	90%	90%	90%	90%	90%	90%	90%
ML-skew	90%	90%	90%	90%	90%	90%	90%	90%	90%
Bootstrap	90%	90%	90%	90%	90%	90%	90%	90%	90%
	Median error								
ML	0.00	0.00	-0.01	-0.01	-0.03	-0.06	-0.10	-0.17	-0.21
ML-skew	0.00	0.00	-0.01	-0.01	-0.03	-0.06	-0.10	-0.17	-0.25
Bootstrap	0.00	0.00	-0.01	-0.01	-0.03	-0.06	-0.13	-0.18	-0.22
	Median absolute error								
ML	0.12	0.26	0.39	0.77	1.65	3.54	7.04	10.64	23.03
ML-skew	0.12	0.26	0.39	0.77	1.65	3.54	7.04	10.64	22.98
Bootstrap	0.12	0.26	0.39	0.77	1.65	3.54	7.03	10.63	23.06

Panel B: Log-normal-with-crashes									
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.25	0.60	0.94	1.98	4.45	9.80	19.70	29.75	63.82
	Coverage rates								
ML	85%	87%	87%	89%	90%	91%	92%	92%	93%
ML-skew	85%	87%	87%	89%	90%	91%	92%	92%	93%
Bootstrap	85%	87%	87%	89%	90%	91%	92%	92%	93%
	Median error								
ML	0.03	0.03	0.02	0.00	-0.06	-0.07	0.29	0.94	5.09
ML-skew	0.00	0.01	0.01	0.02	0.03	0.03	0.06	0.05	-0.34
Bootstrap	0.00	0.01	0.01	0.02	0.02	0.01	0.00	0.01	-0.10
	Median absolute error								
ML	0.11	0.25	0.38	0.78	1.69	3.60	7.13	10.65	22.93
ML-skew	0.11	0.25	0.38	0.78	1.68	3.61	7.10	10.64	22.29
Bootstrap	0.11	0.25	0.38	0.77	1.69	3.61	7.13	10.61	22.27

Panel C: SV									
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.26	0.64	1.00	2.10	4.66	10.10	19.84	29.49	61.13
	Coverage rates								
ML	85%	87%	88%	89%	90%	90%	91%	91%	92%
ML-skew	85%	87%	88%	89%	90%	90%	91%	91%	92%
Bootstrap	86%	87%	88%	89%	90%	90%	91%	91%	92%
	Median error								
ML	0.06	0.06	0.06	0.01	-0.09	-0.24	-0.05	0.47	4.02
ML-skew	0.05	0.06	0.05	0.02	-0.06	-0.21	-0.12	0.23	2.59
Bootstrap	0.05	0.06	0.05	0.02	-0.07	-0.20	-0.13	0.18	2.71
	Median absolute error								
ML	0.12	0.26	0.40	0.81	1.75	3.72	7.20	10.65	22.03
ML-skew	0.12	0.26	0.40	0.81	1.75	3.72	7.20	10.65	22.00
Bootstrap	0.12	0.26	0.40	0.81	1.75	3.72	7.21	10.62	22.06

Panel D: Long-term reversals									
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.43	0.85	1.23	2.29	4.55	9.04	16.78	24.29	48.62
	Coverage rates								
ML	90%	92%	92%	93%	93%	93%	92%	92%	90%
ML-skew	90%	92%	92%	93%	93%	93%	92%	92%	90%
Bootstrap	90%	92%	92%	93%	93%	93%	92%	92%	90%
	Median error								
ML	-0.10	-0.14	-0.17	-0.16	0.00	0.72	2.62	4.94	14.52
ML-skew	-0.10	-0.14	-0.17	-0.16	0.00	0.72	2.62	4.93	14.57
Bootstrap	-0.10	-0.14	-0.17	-0.16	0.00	0.71	2.58	4.88	14.51
	Median absolute error								
ML	0.15	0.29	0.41	0.76	1.53	3.08	5.96	8.99	19.69
ML-skew	0.15	0.29	0.41	0.75	1.53	3.08	5.96	8.99	19.68
Bootstrap	0.15	0.29	0.41	0.75	1.52	3.08	5.95	9.00	19.67

Table 3: Descriptive statistics for individual country returns.

The table shows descriptive statistics for the annual returns of the 21 countries with a full history in the DMS data set. For each country, there are 121 annual return observations, spanning the period from 1900 to 2020. The final row gives the corresponding statistics based on the pooled sample from all 21 countries, containing a total of $2,541$ observations. The first three columns gives, in percent, the arithmetic mean ($\bar{r}_a$), the geometric mean ($\bar{r}_g$), and the standard deviation (i.e., volatility) of the net returns. The following two columns provide the skewness and kurtosis, and the final two columns indicate (in percent) the minimum and maximum annual net returns.

	$\bar{r}_a$ (%)	$\bar{r}_g$ (%)	Stdev (%)	Skew	Kurt	Min (%)	Max (%)
Australia	8.27	6.78	17.36	-0.24	3.22	-42.51	51.48
Austria	5.02	0.86	30.32	1.25	7.12	-59.58	132.75
Belgium	5.30	2.68	23.49	0.52	4.52	-48.90	105.08
Canada	7.05	5.71	16.73	0.01	2.91	-33.77	55.20
Denmark	7.60	5.75	20.64	1.27	7.86	-49.17	107.81
Finland	9.31	5.55	29.26	1.20	8.17	-61.47	161.72
France	5.80	3.35	22.73	0.32	2.78	-41.48	66.07
Germany	8.08	3.33	31.06	1.44	8.79	-90.77	154.60
Ireland	6.93	4.36	22.70	0.23	3.87	-65.42	68.39
Italy	5.94	2.11	28.05	0.71	5.41	-72.85	120.66
Japan	8.67	4.24	28.95	0.50	5.34	-85.51	121.08
Netherlands	7.11	5.10	20.98	0.82	6.01	-50.43	101.59
New Zealand	8.13	6.48	19.07	1.12	9.70	-54.74	105.31
Norway	7.16	4.35	26.26	2.12	13.73	-53.61	166.89
Portugal	8.48	3.72	33.64	1.62	8.28	-76.60	151.83
South Africa	9.14	7.06	21.73	0.90	5.58	-52.23	102.88
Spain	5.60	3.46	21.55	0.72	4.85	-43.32	99.42
Sweden	8.14	6.05	20.84	0.08	3.22	-42.52	67.53
Switzerland	6.37	4.61	19.24	0.32	3.34	-37.83	59.36
UK	7.18	5.39	19.50	0.65	6.76	-56.60	99.31
U.S.	8.54	6.60	19.78	-0.23	2.75	-38.57	55.84
Full Panel	7.32	4.63	24.03	1.02	8.15	-90.77	166.89

Table 4: Empirical estimates of global long-run returns

The table shows the pooled point estimates and 90 percent confidence intervals (in parentheses) for the long-run distributions of the global gross returns based on the panel of 21 countries with a full history in the DMS data set. Panels A and B show results for 10-year and 30-year compounding horizons, respectively. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile). Results are presented using data for the entire sample period (Full sample) and for a subsample starting in 1960 (Post-1960), as indicated in the row headers. For each sample and horizon, results for the pooled MLE, the pooled skewness-corrected MLE (ML-skew), and the pooled FF bootstrap estimator are presented.

	Panel A: $T = 10$						
	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
	I. ML						
Full sample	0.47	0.61	0.96	1.57	2.58	4.03	5.26
	(0.35-0.63)	(0.47-0.81)	(0.75-1.23)	(1.25-1.98)	(2.08-3.20)	(3.28-4.95)	(4.29-6.45)
Post-1960	0.50	0.66	1.03	1.71	2.84	4.47	5.87
	(0.31-0.81)	(0.42-1.03)	(0.68-1.56)	(1.18-2.49)	(2.01-4.01)	(3.22-6.20)	(4.26-8.09)
	II. ML-skew						
Full sample	0.44	0.60	0.98	1.64	2.64	3.92	4.88
	(0.33-0.58)	(0.45-0.78)	(0.76-1.26)	(1.31-2.07)	(2.13-3.27)	(3.19-4.81)	(3.98-5.98)
Post-1960	0.48	0.65	1.05	1.75	2.87	4.41	5.67
	(0.30-0.78)	(0.41-1.02)	(0.69-1.58)	(1.20-2.54)	(2.03-4.06)	(3.18-6.13)	(4.11-7.82)
	III. FF bootstrap						
Full sample	0.46	0.63	1.00	1.62	2.56	3.84	4.93
	(0.34-0.61)	(0.48-0.83)	(0.78-1.29)	(1.29-2.04)	(2.06-3.17)	(3.13-4.72)	(4.02-6.05)
Post-1960	0.49	0.66	1.06	1.74	2.83	4.36	5.66
	(0.30-0.78)	(0.42-1.03)	(0.70-1.60)	(1.20-2.54)	(2.00-4.00)	(3.14-6.05)	(4.11-7.80)

	Panel B: $T = 30$						
	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
	I. ML						
Full sample	0.48	0.76	1.65	3.89	9.17	19.85	31.51
	(0.22-1.05)	(0.36-1.63)	(0.80-3.40)	(1.96-7.74)	(4.75-17.73)	(10.48-37.59)	(16.80-59.10)
Post-1960	0.60	0.96	2.10	5.03	12.06	26.50	42.43
	(0.16-2.17)	(0.27-3.34)	(0.64-6.90)	(1.63-15.53)	(4.13-35.24)	(9.46-74.24)	(15.47-116.43)
	II. ML-skew						
Full sample	0.45	0.74	1.69	4.07	9.40	19.29	29.21
	(0.20-0.97)	(0.35-1.58)	(0.82-3.48)	(2.04-8.09)	(4.86-18.17)	(10.19-36.54)	(15.57-54.79)
Post-1960	0.58	0.94	2.12	5.13	12.19	26.16	41.02
	(0.16-2.10)	(0.27-3.30)	(0.65-6.97)	(1.66-15.84)	(4.17-35.62)	(9.34-73.30)	(14.95-112.55)
	III. FF bootstrap						
Full sample	0.45	0.76	1.72	4.04	9.16	18.94	29.28
	(0.21-0.99)	(0.36-1.62)	(0.83-3.54)	(2.03-8.03)	(4.74-17.70)	(10.00-35.88)	(15.61-54.93)
Post-1960	0.58	0.95	2.13	5.11	12.06	25.87	40.80
	(0.16-2.09)	(0.27-3.31)	(0.65-6.99)	(1.66-15.78)	(4.13-35.22)	(9.23-72.48)	(14.87-111.95)

Figure 1: Simulation results for $T = 120$.

The figure shows simulation results based on estimates with a sample size of $n = 1,440$ and a compounding horizon of $T = 120$. Monthly period returns are generated according to the four different return models described in the main text: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals (Panel D). All four specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. Three different estimators are considered: (i) the MLE, (ii) the skewness-corrected MLE (ML-skew), and (iii) the FF bootstrap estimator. The results are based on $10,000$ samples. Each panel shows the median and the 5^{th} and 95^{th} percentiles of the estimated quantiles of the long-run gross return distributions (calculated across the $10,000$ estimates obtained from the simulated samples), for each of the three estimation procedures. The solid line shows the true (population) quantiles in each graph. The dotted line shows the median estimates for the MLE and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the ML estimates. The dashed lines show the median and the 5^{th} and 95^{th} percentiles of the skewness-corrected ML estimates of each quantile. The dashed-and-dotted lines show the corresponding estimates for the bootstrap estimator.

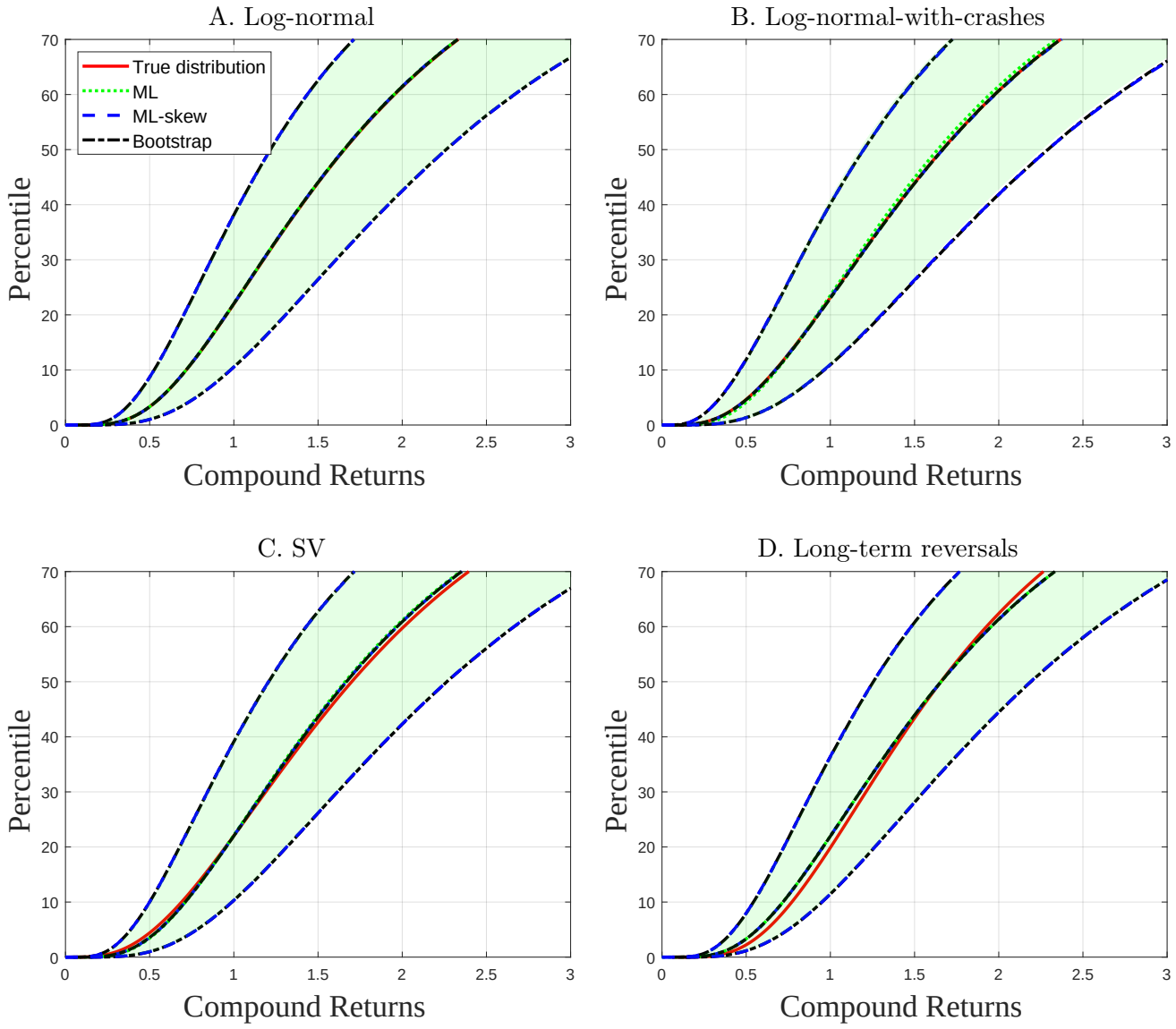


Figure 2: Simulation results for $T = 360$.

The figure shows simulation results based on estimates with a sample size of $n = 1,440$ and a compounding horizon of $T = 360$. Monthly period returns are generated according to the four different return models described in the main text: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals (Panel D). All four specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. Three different estimators are considered: (i) the MLE, (ii) the skewness-corrected MLE (ML-skew), and (iii) the FF bootstrap estimator. The results are based on $10,000$ samples. Each panel shows the median and the 5^{th} and 95^{th} percentiles of the estimated quantiles of the long-run gross return distributions (calculated across the $10,000$ estimates obtained from the simulated samples), for each of the three estimation procedures. The solid line shows the true (population) quantiles in each graph. The dotted line shows the median estimates for the MLE and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the ML estimates. The dashed lines show the median and the 5^{th} and 95^{th} percentiles of the skewness-corrected ML estimates of each quantile. The dashed-and-dotted lines show the corresponding estimates for the bootstrap estimator.

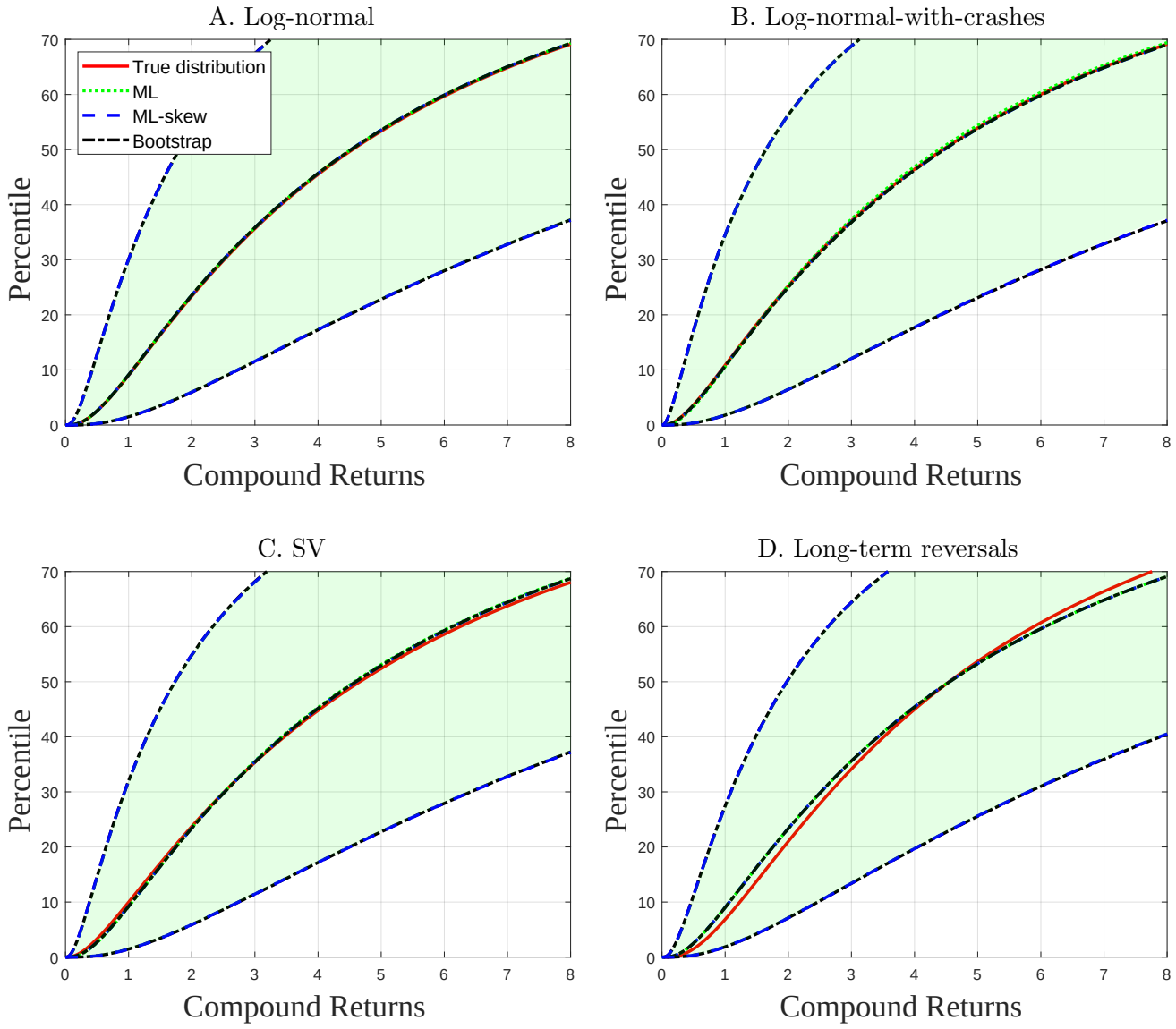


Figure 3: Empirical estimates of global long-run returns

The figure shows the pooled point estimates and 90 percent confidence intervals for the long-run distributions of the global gross returns, based on the panel of 21 countries with a full history in the DMS data set. Panels A1 and A2 show results for a 10-year compounding horizon, and Panels B1 and B2 show results for a 30-year compounding horizon. Results are presented using data for the entire sample period (Full sample, Panels A1 and B1) and for a subsample starting in 1960 (Post-1960 sample, Panels A2 and B2). For each sample and horizon, results for the pooled MLE, the pooled skewness-corrected MLE (ML-skew), and the pooled FF bootstrap estimator are presented. Each panel shows the point estimates and corresponding 90% confidence bands for the long-run return distributions, for each of the three estimation procedures. The dotted line and the shaded area show point estimates and confidence bands using the pooled MLE. The dashed lines show the point estimates (the middle of the three dashed lines) and the confidence bands using the pooled skewness-corrected MLE. The dashed-and-dotted lines show the point estimates (the middle of the three dotted lines) and the confidence bands using the pooled FF bootstrap estimator.

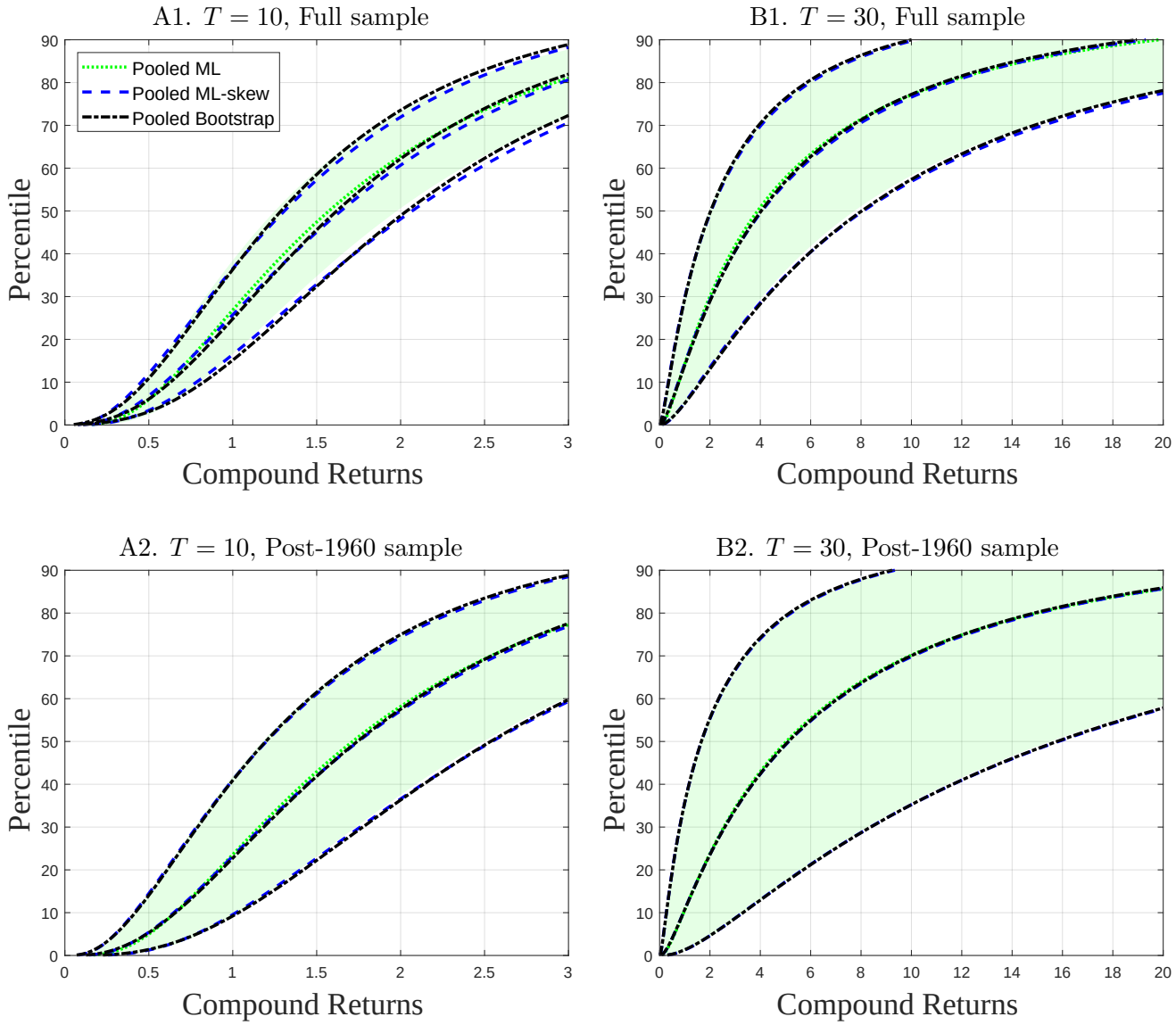


Figure 4: U.S. returns versus global returns

The figure shows estimates of the long-run distributions of the global gross returns along with the corresponding estimates for the U.S. gross returns. Panels A1 and A2 show results for a 10-year compounding horizon, and Panels B1 and B2 show results for a 30-year compounding horizon. Results are presented using data for the entire sample period (Full sample, Panels A1 and B1) and for a subsample starting in 1960 (Post-1960 sample, Panels A2 and B2). All estimates are based on the skewness-corrected MLE. The estimates for the global distribution are formed from the pooled panel of 21 countries with a full history in the DMS data set. The dashed line and the shaded area show point estimates and 90% confidence bands, respectively, for the global return distributions. The solid line shows the point estimates for the long-run return distribution based on returns data for the U.S; the dotted lines show the corresponding 90% confidence bands. In addition, p-values are shown for the test of the null hypothesis that a given percentile of the global return distribution is identical to the corresponding percentile for the U.S. distribution. Specifically, p-values for the 5^{th}, 50^{th} and 95^{th} percentiles, labeled p-val(p_5), p-val(p_{50}), and p-val(p_{95}), respectively, are displayed.

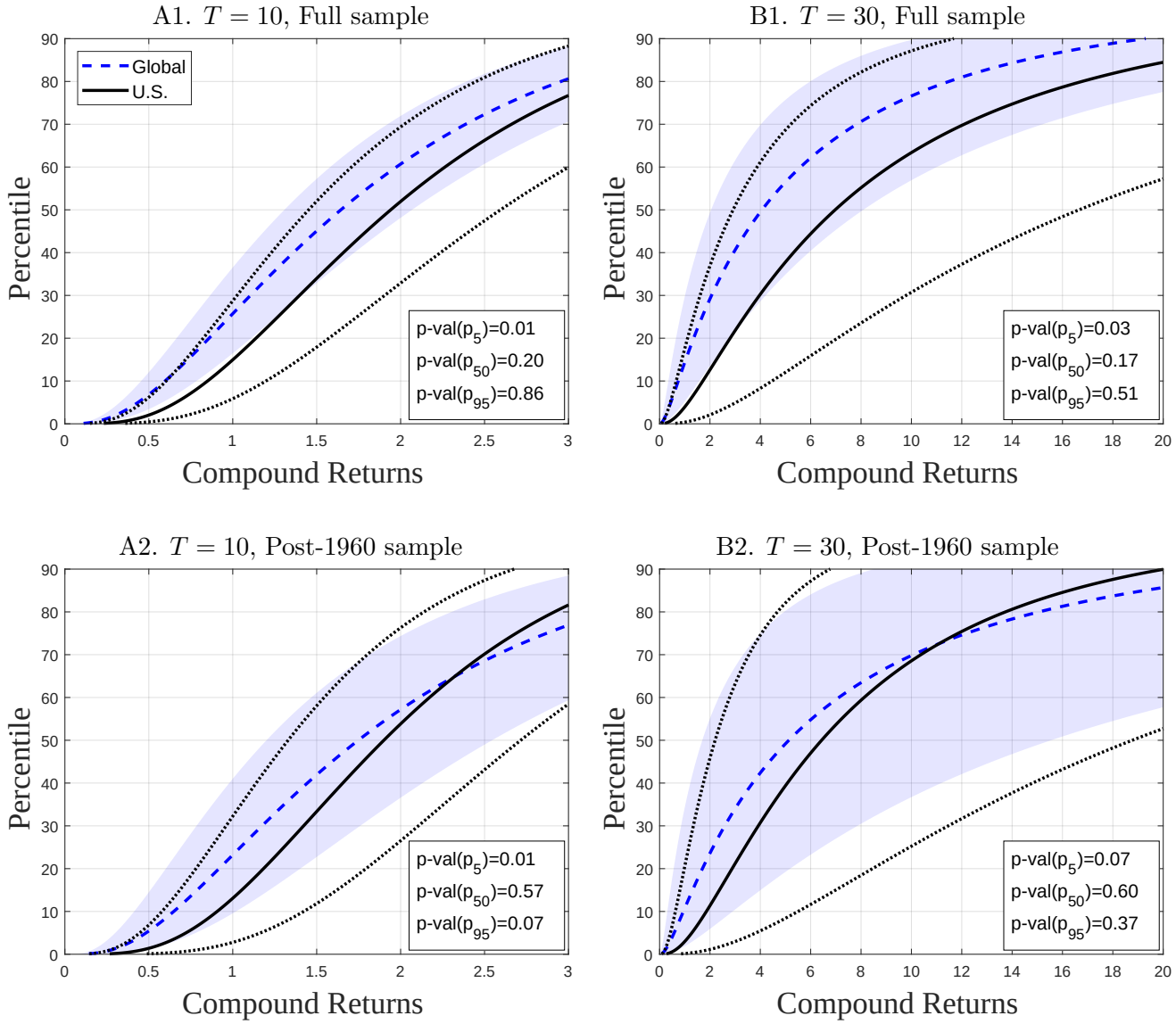


Figure 5: Individual country returns versus global returns - Full sample

The figure shows estimates of selected percentiles of the long-run gross return distribution for individual countries (horizontal lines) along with the corresponding pooled estimates for the global gross returns (vertical lines and shaded areas). All results are based on the full sample period from 1900 to 2020, using the skewness-corrected MLE. In each panel, the horizontal solid lines represent 90% confidence intervals for the given long-run return percentile indicated in the panel header, for a given country. The solid circles on each of these horizontal lines represent the corresponding country-specific point estimate of that return percentile. The solid vertical line and the shaded area show the corresponding point estimate and 90% confidence interval for the global long-run returns. These are formed from the pooled panel of 21 countries with a full history in the DMS data set. In addition, the p-value for the test of the null hypothesis that the given percentile of the global and country-specific return distributions is identical is shown for each country. Significance at the 5% and 10% levels, according to Holm's (1979) multiple testing procedure, is indicated with a dagger (†) and double-dagger (‡), respectively, next to the p-values. The left-hand side panels (A1-A3) show results for 10-year compounding horizons and the right-hand side panels (B1-B3) show results for 30-year compounding horizons. The top panels (A1 and B1) show results for the 5^{th} percentile of the long-run return distribution; the middle panels (A2 and B2) show results for the 50^{th} percentile of the long-run return distribution; the bottom panels (A3 and B3) show results for the 95^{th} percentile of the long-run return distribution.

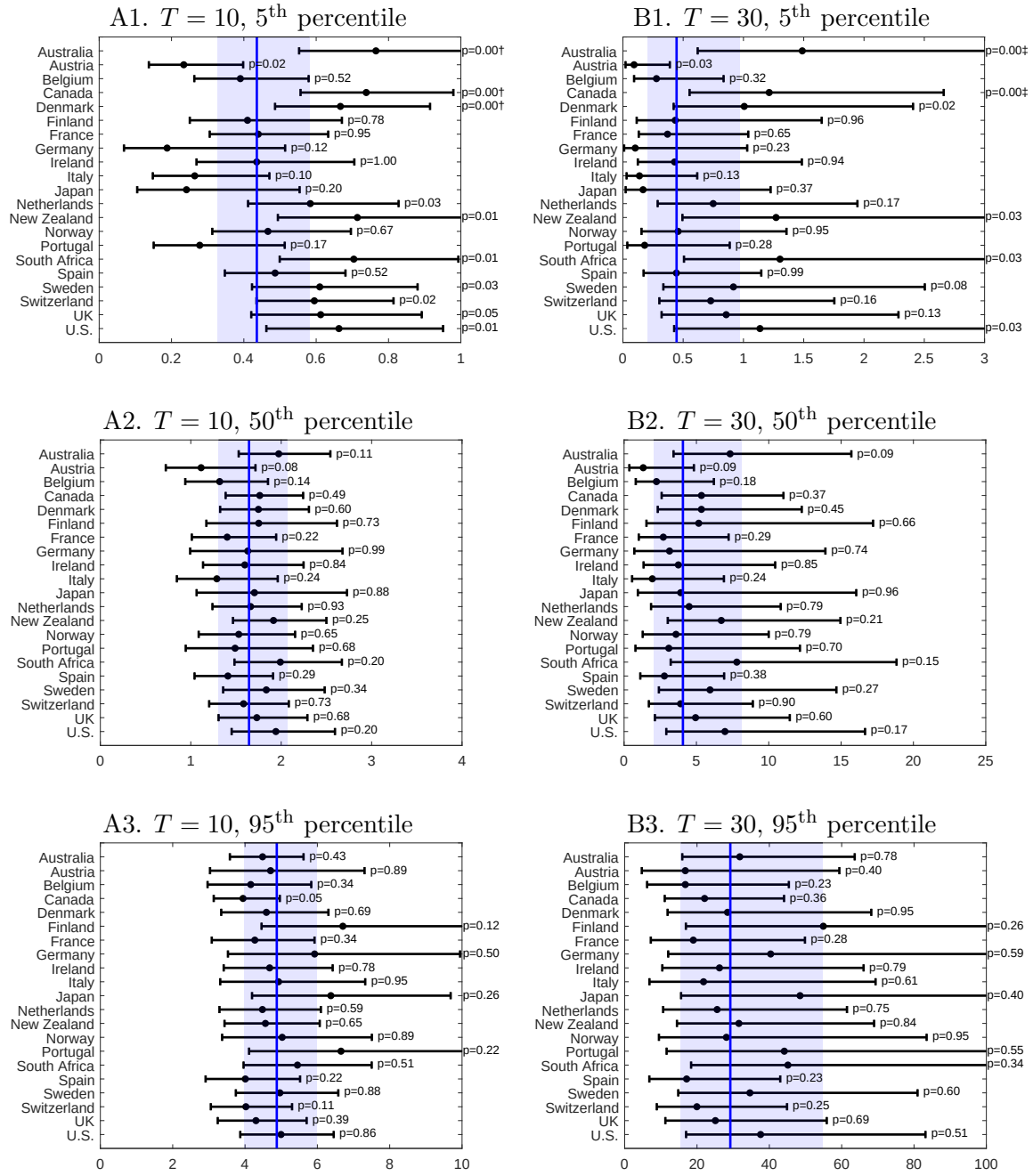
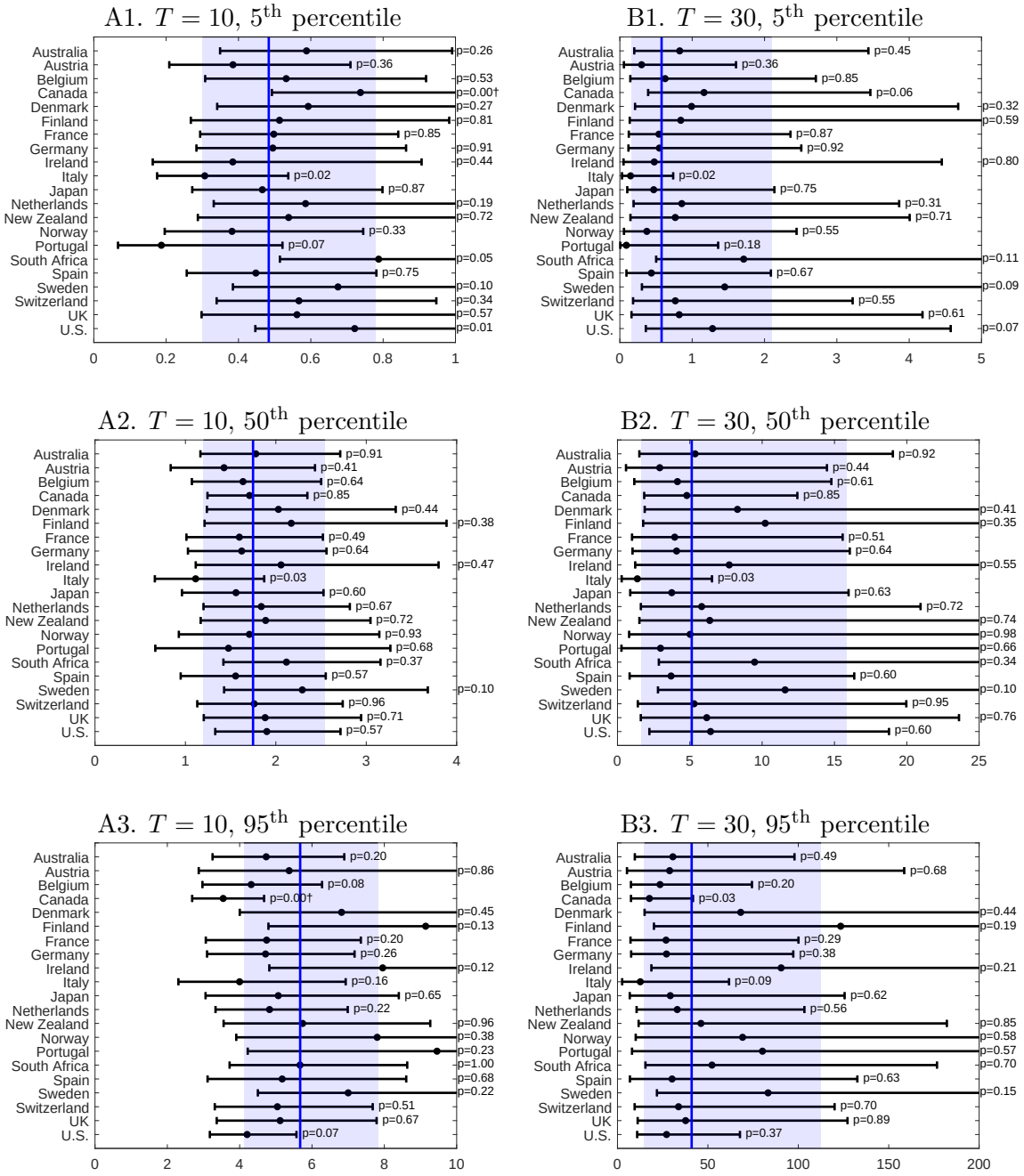


Figure 6: Individual country returns versus global returns - Post-1960 sample

The figure shows estimates of selected percentiles of the long-run gross return distribution for individual countries (horizontal lines) along with the corresponding pooled estimates for the global gross returns (vertical lines and shaded areas). All results are based on the subsample starting in 1960 and ending in 2020, using the skewness-corrected MLE. In each panel, the horizontal solid lines represent 90% confidence intervals for the given long-run return percentile indicated in the panel header, for a given country. The solid circles on each of these horizontal lines represent the corresponding country-specific point estimate of that return percentile. The solid vertical line and the shaded area show the corresponding point estimate and 90% confidence interval for the global long-run returns. These are formed from the pooled panel of 21 countries with a full history in the DMS data set. In addition, the p-value for the test of the null hypothesis that the given percentile of the global and country-specific return distributions is identical is shown for each country. Significance at the 5% and 10% levels, according to Holm's (1979) multiple testing procedure, is indicated with a dagger (†) and double-dagger (‡), respectively, next to the p-values. The left-hand side panels (A1-A3) show results for 10-year compounding horizons and the right-hand side panels (B1-B3) show results for 30-year compounding horizons. The top panels (A1 and B1) show results for the 5^{th} percentile of the long-run return distribution; the middle panels (A2 and B2) show results for the 50^{th} percentile of the long-run return distribution; the bottom panels (A3 and B3) show results for the 95^{th} percentile of the long-run return distribution.



Online Appendix for "Long-Run Stock Return Distributions: Empirical Inference and Uncertainty"

Andreas Dzemski Adam Farago Erik Hjalmarsson
Tamas Kiss

October 14, 2025

This internet appendix covers the following items:

A1. Additional simulation results, showing the impact of altering the sample size in the time-series case, altering the sampling frequency (e.g., annual instead of monthly), as well as results for the panel data case. Simulation results for "direct" estimation are also shown.

A2. Details on the return specifications used in the simulations.

A3. A discussion of the ACO block bootstrap procedure.

A4. Additional empirical results.

A5. Formal results for estimation and testing with panel data.

A6. Formal results for consistency of point estimates of quantiles of gross returns.

A7. Proofs of formal theoretical results.

A1 Additional simulation results

A1.1 Smaller sample size

We repeat the same simulation exercise as in the main text, but using a sample with $n = 720$ monthly observations, rather than $n = 1,440$. That is, 60 years worth of data are used, rather than 120 years. The results are shown in Figures A1 and A2, which follow exactly the same format as Figures 1 and 2 in the main text.

As one would expect, the precision of the estimates is reduced substantially. The point estimates of the 30-year distribution (Figure A2) are now almost completely uninformative.

While many markets have data available from around 1900, it is not obvious how informative the earlier observations are for today's (or tomorrow's) conditions. A sample length of 60 years might therefore be more relevant for many practical purposes. Fama and French (2018, FF) restrict their analysis to post-1963 data because of such considerations. The case with $n = 1,440$ might be viewed as the very best we could hope for—120 years of informative data—but real-world situations are likely to often be less favorable.

A1.2 Different sampling frequencies

As discussed in detail in the main text, the primary source of uncertainty in the estimates of the long-run quantiles is the mean of (log) period returns. It is well known that the precision of the estimation of the mean is not affected by the sampling frequency (e.g., Merton, 1980). There is therefore no reason to expect any significant precision gains from using data sampled at higher frequency, such as daily, when estimating the long-run quantiles. Vice versa, there is also no reason to expect any decrease in precision when using lower frequency data, such as annual. In this subsection, we verify in simulations that these conclusions hold.

We simulate daily, monthly, and annual data from the log-normal and stochastic volatility (SV) specifications. In each case, the returns are parameterized such that the corresponding monthly returns have the same mean and volatility as in the simulations in the main text (i.e., mean $\mu = 1.006$ and volatility $\sigma = 0.06$). We focus on the log-normal

and SV returns since the monthly specifications trivially map into corresponding daily specifications in these cases. The annual returns are simply generated as non-overlapping 12-month returns, obtained from the simulated monthly data. The sample size is set to 120 years, such that $30,240$ daily observations (each month has 21 trading days), $1,440$ monthly observations, and 120 annual observations are used in each simulated sample. $10,000$ simulated samples are generated in each case.

The results are shown in Figure A3. The figure follows the same format as previous simulation figures, but results for 10-year compounding horizons are now shown in the left panels (A1 and A2) and results for 30-year compounding horizons are shown in the right panels (B1 and B2). We show results only for the skewness-corrected MLE; the results for the MLE and the FF bootstrap estimator are very similar. As is immediately seen from the figures, there is very little difference between the estimates using data sampled at the three different frequencies. In the log-normal return specification, there is hardly any noticeable difference between the three sets of estimates. In the SV specification, there are some very small visible differences across the three sampling frequencies. While it is tempting to think that the daily estimates must be more precise, a close inspection shows that the median *annual* estimates actually track the true distribution marginally better than the daily estimates; the monthly estimates tend to be in between. The spread in the estimates, as measured by the reported 5^{th} and 95^{th} percentiles of the estimates, are also very similar for the three sampling frequencies. Importantly, compared to the overall estimation uncertainty, these differences are all tiny.

A1.3 "Direct" estimation results

While the precision of the mean estimator of the log returns is not affected by the sampling frequency, the precision of the volatility estimator *is* affected. As documented in Figure A3, this change in volatility precision has no practical impact on the estimated long-run distribution, when comparing daily, monthly, or annual sampling frequencies. However, if the sampling frequency is coarse enough to result in very poor volatility estimates, the resulting estimator of the long-run distribution will eventually be affected.

In the extreme case, one follows the "direct" estimation approach and uses only the non-overlapping observations of long-run returns for inference, as discussed in Section 2 of the main text. That is, with $n = 1,440$ monthly observations, the direct approach would simply form estimates based on the $n/120 = 12$ observations of 10-year returns or the $n/360 = 4$ observations of 30-year returns.

In Figure A4, we show simulation results from using "direct" inference. The simulation specification is the same as in the previous subsection, where we compared different sampling frequencies. We consider "direct" estimation, which can simply be viewed as an approach where we sample the data at the same frequency as the investment horizons that we are interested in. I.e., we sample the data at the 10-year or 30-year horizon, resulting in 12 and 4 return observations, respectively. We focus on the normal MLE, which only requires the mean and variance as inputs. As a comparison to the "direct" estimates, we show results for the ("indirect") MLE, based on monthly data.

The layout of Figure A4 is the same as for Figure A3, except we now show results for the MLE rather than the skewness-corrected MLE. The shaded area in each panel show the 5^{th} and 95^{th} percentiles for the (indirect) monthly ML estimates. The dashed-and-dotted lines show the median, as well as the 5^{th} and 95^{th} percentiles, of the direct estimates. Starting with the 10-year horizon (Panels A1 and A2), where the direct estimator is based on 12 observations, we see that the differences between the direct estimator and the monthly indirect estimator are not that dramatic. In Panel A1 (and B1), the standard (indirect) MLE is optimal. The direct estimator clearly performs somewhat worse, but not terribly so. At the 30-year horizon, the differences are more stark. The direct estimator now only uses 4 observations, and the resulting lack of precision in the volatility estimator has a big impact on the resulting long-run distribution. The indirect estimator clearly dominates, both in terms of precision and bias.

A1.4 Panel data results

We next present simulation results for estimates based on a panel data set, rather than a single time series.

A1.4.1 Return distributions

Suppose there are log returns $y_{t,i}$, $i = 1, .., K$, representing K different market indexes. These are generated according to a 1-factor model,

$$y_{t,i} = \beta_i z_t + \epsilon_{t,i}, \tag{A1}$$

where z_t and $\epsilon_{t,i}$ are independent. In each generated sample, the β_is are drawn from a uniform distribution over $[0.7, 1.3]$, such that the β_is range from 0.7 to 1.3 and are on average equal to 1. The number of markets is set to $K = 20$ and the number of monthly observations is set to $n = 1,440$. Each panel thus contain $28,800$ monthly observations. The common factor z_t captures 40% of the total variation in $y_{t,i}$ ($\lambda = 0.4$ in the notation of Section 4 in the main text), which is very similar to the value we find empirically in Section 6 in the main text. According to the derivations in Section 4.2 in the main text, the effective sample size is therefore equivalent to about $3,350$ months of independent time-series observations.

We consider two different specifications.

- *Log-normal panel:* z_t and $\epsilon_{t,i}$ are both i.i.d. normally distributed across time, and $\epsilon_{t,i}$ is mean zero. The gross returns $x_{t,i}$ are therefore log-normally distributed.

- *Log-normal-with-crashes panel:* z_t and $\epsilon_{t,i}$ are both drawn from mixed-normal distributions with crashes. The mean of z_t is adjusted to achieve a specific mean return, controlling for the mean effect of the crashes. The gross returns follow a log-normal-with-crashes distribution.

These two specifications are the panel data analogues of the corresponding time-series specifications in the main text. In the pure log-normal specification, the extension to a panel framework is completely straightforward. In the crash model, crashes can occur in both the systematic and the idiosyncratic part. To keep the specification as aligned as possible with the one in the time-series case, we assume that crashes occur with a probability of $p = 1/100$ in a given period and that a given crash results in an expected

loss of 30%. However, since crashes can occur in either the systematic or the idiosyncratic part, we now assume that there is a $p/2 = 1/200$ probability that a crash occurs in either of these components.[1] Additional details on the panel specifications are given in Online Appendix A2.

Since the panel is heterogeneous in the sense that β_i varies across i, the "population" distribution is not uniquely defined. In the below simulation results, we treat the distribution of the average, or typical, asset in the panel as the population distribution; i.e., the distribution for an asset with $\beta_i = 1$. As in the time-series simulations, we parameterize the models such that $\mu = 1.006$ and $\sigma = 0.06$ for the $\beta_i = 1$ asset.

A1.4.2 Estimates based on panel data

We consider the pooled MLE, as well as the pooled skewness-corrected MLE and the pooled bootstrap estimator, as defined in Section 4.1 of the main text. Confidence intervals are calculated from equations (35) and (36) in the main text, but with the intervals centered on each of the three pooled estimators.[2] All results are based on $10,000$ repetitions; i.e., $10,000$ panels of data are simulated for each specification.

The results are shown in Table A1 and Figure A5. In both the table and the figure, the top panels correspond to the log-normal return specification and the bottom panels to the log-normal-with-crashes return specification. The left-hand side shows the case with $T = 120$ and right-hand side panels the case with $T = 360$. The panels in Figure A5 follow the same format as Figures 1 and 2 in the main text. The plain solid line shows the true quantiles in each case, defined as the population distribution for a $\beta_i = 1$ asset.

A comparison of the results in Figure A5 with the corresponding time-series results in Figures 1 and 2 in the main text highlights that the panel estimates have higher precision, but not dramatically so. In Panel A in Figure 2 in the main text, the time-series estimates

[1] The probability of exactly one crash is no longer exactly equal to 1%, but the approximation is very close. Two crashes can in theory occur in a single period, but the probability of this happening is tiny (0.0025%).

[2] As pointed out in the main text, in the panel case the confidence intervals are only formally justified for the MLE and the skewness-corrected MLE, but we also apply them to the pooled FF bootstrap estimator. As seen in the simulation results, the pooled skewness-corrected MLE and the pooled FF bootstrap estimator behave almost identically.

of the 10^{th} percentile of the 30-year distribution stretches from 0.4 to around 2.7. In Panel B1 of Figure A5, the corresponding range for the panel estimates are 0.6 to 2.0. Apart from this increase in precision, the panel estimators behave very similarly to the corresponding time-series cases.

The actual coverage rates of the confidence intervals (see Table A1) are in most cases close to the nominal coverage rate of 90 percent. For the log-normal specification (Panels A1 and A2), the actual coverage rates are all between 88% and 90%, across all quantiles and all three pooled estimators. In the log-normal-with-crashes specification (Panels B1 and B2), the coverage rates for the confidence intervals around the skewness-corrected ML estimator and the FF boostrap estimator are all above 85% in the $T = 120$ case and above 87% in the $T = 360$. As remarked upon in the discussion of the simulation results in the main text, the coverage rates of the confidence intervals are expected to improve as T increases.

A2 Details on simulation specifications

A2.1 Time-series specifications

A2.1.1 Log-normal-with-crashes

The i.i.d. log returns y_t can be written as $y_t = \tilde{y}_t + \kappa b_t$, where $\tilde{y}_t \sim N\left(\tilde{\mu}_y, \tilde{\sigma}_y^2\right)$, $b_t \sim Bernoulli\,(p)$, and κ is a constant. That is, b_t is a discrete random variable that takes on a value of 1 with probability p and 0 with probability $1 - p$. $\tilde{y}_t$ and b_t are both i.i.d. across time and b_t is independent of $\tilde{y}_t$. The distribution of y_t is a mixture of two normals and the gross-returns is a mixture of two log-normals,

$$x_t \sim \begin{cases} LN\left(\tilde{\mu}_y, \tilde{\sigma}_y^2\right) \text{ with probability } 1-p \\ LN\left(\kappa + \tilde{\mu}_y, \tilde{\sigma}_y^2\right) \text{ with probability } p \end{cases}. \tag{A2}$$

If p is small, $\kappa < 0$, and $|\kappa| >> \tilde{\sigma}_y$, the p-probability outcome corresponds to a (low-probability) crash. The standard log-normal distribution is a special case where $p = 0$

and/or $\kappa = 0$. With $\tilde{y}_t$ and b_t independent, it follows easily that the first two moments of x_t are given by

$$\mathbb{E}[x_t] = (1 - p + pe^{\kappa}) e^{\tilde{\mu}_y + \frac{\tilde{\sigma}_y^2}{2}} = (1-p) e^{\tilde{\mu}_y + \frac{\tilde{\sigma}_y^2}{2}} + pe^{\kappa + \tilde{\mu}_y + \frac{\tilde{\sigma}_y^2}{2}}, \tag{A3}$$

$$\mathbb{E}\left[x_t^2\right] = \left(1 - p + pe^{2\kappa}\right) e^{2\tilde{\mu}_y + 2\tilde{\sigma}_y^2}, \tag{A4}$$

and the variance is subsequently equal to

$$Var(x) = \left(1 - p + pe^{2\kappa}\right) e^{2\tilde{\mu}_y + 2\tilde{\sigma}_y^2} - (1 - p + pe^{\kappa})^2 e^{2\tilde{\mu}_y + \tilde{\sigma}_y^2}. \tag{A5}$$

Conditional on a crash occurring,

$$\mathbb{E}[x_t | b_t = 1] = e^{\kappa + \tilde{\mu}_y + \frac{\tilde{\sigma}_y^2}{2}} \approx e^{\kappa}, \tag{A6}$$

where the last approximation is fairly accurate for any large crash. For a given crash-probability p and crash-size κ, $\tilde{\mu}_y$ and $\tilde{\sigma}_y^2$ can be set to match the mean, μ, and variance, σ^2, of the gross returns by using,

$$\tilde{\mu}_y = \log\left(\frac{\mu}{(1 - p + pe^{\kappa})}\right) - \frac{\tilde{\sigma}_y^2}{2}, \tag{A7}$$

$$\tilde{\sigma}_y^2 = \log\left(\left(\frac{\sigma^2}{\mu^2} + 1\right) \times \frac{(1 - p + pe^{\kappa})^2}{(1 - p + pe^{2\kappa})}\right). \tag{A8}$$

In the simulated model, $p = 1/100$ and $e^{\kappa} = 0.7$, such that a 30% loss (in expectation) occurs on average once every 100 months (≈8 years). With $\mu = 1.006$ and $\sigma = 0.06$, this gives $\tilde{\mu}_y = 0.00766$ and $\tilde{\sigma}_y^2 = 0.00265$, which implies that

$$\mathbb{E}[x_t | b_t = 1] = e^{\kappa + \tilde{\mu}_y + \frac{\tilde{\sigma}_y^2}{2}} = 0.7 e^{0.00766 + \frac{0.00265}{2}} = 0.7063, \tag{A9}$$

$$\mathbb{E}[x_t | b_t = 0] = e^{\tilde{\mu}_y + \frac{\tilde{\sigma}_y^2}{2}} = e^{0.00766 + \frac{0.00265}{2}} = 1.009. \tag{A10}$$

That is, to achieve the same expected return as in a process without crashes, the mean in non-crash periods needs to be increased from 1.006 to 1.009.

A2.1.2 Long-term reversals

The log returns y_t follow a moving average process of order q (MA(q)), driven by i.i.d. normal innovations,

$$y_t = \mu_y + \sum_{k=0}^{q} \theta_k u_{t-k}, \quad \theta_0 = 1, \tag{A11}$$

$$u_t \sim i.i.d.N\left(0, \sigma_u^2\right). \tag{A12}$$

The variance of y_t is given by

$$\sigma_y^2 = Var\left(y_t\right) = \sum_{k=0}^{q} Var\left(\theta_k u_{t-k}\right) = \sigma_u^2 \sum_{k=0}^{q} \theta_k^2. \tag{A13}$$

The *long-run* variance of y_t equals

$$Lr.Var\left(y_t\right) = \sum_{j=-\infty}^{\infty} Cov\left(y_t, y_{t+j}\right) = \sigma_u^2 \left(\sum_{k=0}^{q} \theta_k\right)^2. \tag{A14}$$

The variance ratio of the long-run to short-run variance is given by

$$VR_{LR/SR} = \frac{Lr.Var\left(y_t\right)}{Var\left(y_t\right)} = \frac{\left(\sum_{k=0}^{q} \theta_k\right)^2}{\sum_{k=0}^{q} \theta_k^2}. \tag{A15}$$

The variance-ratio provides a summary of the degree of serial correlation in y_t and we parameterize the process to achieve a certain variance ratio. Specifically, we assume that the MA coefficients are declining for greater lags, and use the parametric form $\theta_k = \frac{\theta_1}{\sqrt{k}}$, $\theta_0 = 1$. In this case,

$$VR_{LR/SR} = \frac{\left(\sum_{k=0}^{q} \theta_k\right)^2}{\sum_{k=0}^{q} \theta_k^2} = \frac{\left(1 + \sum_{k=1}^{q} \frac{\theta_1}{\sqrt{k}}\right)^2}{1 + \sum_{k=1}^{q} \left(\frac{\theta_1}{\sqrt{k}}\right)^2}. \tag{A16}$$

We consider two lag lengths q, equal to 60 and 120 (where the latter one is only used in the simulations in Online Appendix A3), and we target a variance ratio of 0.8. Setting $\theta_1 = -0.0075$ for $q = 60$ and $\theta_1 = -0.0050$ for $q = 120$ almost exactly achieves this variance ratio.

To complete the parameterization of the process, we target a mean μ and variance σ^2 for the gross returns. Since y_t is normally distributed (it is the sum of i.i.d. normal innovations), x_t is log-normal. We therefore specify (μ, σ^2) and calculate $\left(\mu_y, \sigma_y^2\right)$ based on the standard log-normal formula. Further,

$$\sigma_y^2 = Var\left(y_t\right) = \sigma_u^2\left(1 + \sum_{k=0}^{q}\theta_k^2\right) = \sigma_u^2\left(1 + \sum_{k=1}^{q}\left(\frac{\theta_1}{\sqrt{k}}\right)^2\right), \tag{A17}$$

and

$$\sigma_u^2 = \frac{\sigma_y^2}{1 + \sum_{k=1}^{q}\left(\frac{\theta_1}{\sqrt{k}}\right)^2}. \tag{A18}$$

In practice, this adjustment to the variance of the innovations is very small.

A2.2 Panel specifications

The data are generated according to a 1-factor model,

$$y_{t,i} = \beta_i z_t + \epsilon_{t,i}. \tag{A19}$$

A2.2.1 Log-normal

The components z_t and $\epsilon_{t,i}$ are i.i.d. log-normal across time and independent of each other. The variances of z_t and $\epsilon_{t,i}$ are set such that

$$Var\left(z_t\right) = \sigma_z^2 = \lambda \times \sigma_y^2, \tag{A20}$$

and

$$Var\left(\epsilon_{t,i}\right) = \sigma_\epsilon^2 = (1 - \lambda) \times \sigma_y^2. \tag{A21}$$

It follows that

$$Var\left(y_{t,i}\right) = \beta_i^2\lambda\sigma_y^2 + (1 - \lambda) \times \sigma_y^2, \tag{A22}$$

such that for $\beta_i = 1$, $Var\left(y_{t,i}\right) = \sigma_y^2$.

A2.2.2 Log-normal-with-crashes

In the log-normal-with-crashes specification, z_t and $\epsilon_{t,i}$ are both i.i.d. log-normal mixtures. The common component z_t can be written as $z_t = \tilde{z}_t + \kappa b_t^z$, where $\tilde{z}_t \sim N\left(\tilde{\mu}_z, \tilde{\sigma}_z^2\right)$, $b_t^z \sim Bernoulli\left(p/2\right)$, and κ is a constant. The idiosyncratic component $\epsilon_{t,i}$ can be written as $\epsilon_{t,i} = \tilde{\epsilon}_{t,i} + \kappa b_t^\epsilon$, where $\tilde{\epsilon}_{t,i} \sim N\left(0, \tilde{\sigma}_\epsilon^2\right)$, $b_t^\epsilon \sim Bernoulli\left(p/2\right)$, and κ is a constant. $\{\tilde{z}_t, b_t^z, \tilde{\epsilon}_{t,i}, b_t^\epsilon\}$ are i.i.d. across time and mutually independent of each other.

For a given $\tilde{\sigma}_y^2$ and λ, we parameterize

$$\tilde{\sigma}_z^2 = \lambda \times \tilde{\sigma}_y^2 \quad \text{and} \quad \tilde{\sigma}_\epsilon^2 = (1-\lambda) \times \tilde{\sigma}_y^2. \tag{A23}$$

To match the mean, μ, and variance, σ^2, of the gross returns, we use the same transformation described above for the time-series case. For a given (total) crash-probability p and crash-size κ, $\tilde{\mu}_z$ and $\tilde{\sigma}_y^2$ are set to

$$\tilde{\mu}_z = \log\left(\frac{\mu}{(1-p+pe^\kappa)}\right) - \frac{\tilde{\sigma}_y^2}{2}, \tag{A24}$$

$$\tilde{\sigma}_y^2 = \log\left(\left(\frac{\sigma^2}{\mu^2}+1\right) \times \frac{(1-p+pe^\kappa)^2}{(1-p+pe^{2\kappa})}\right). \tag{A25}$$

This mapping relies on a small approximation, effectively treating the sum of the two independent crash distributions, each with a probability of a crash given by $p/2$, as if it was a single crash distribution with the probability of a crash given by p. The values of $\tilde{\mu}_z$ and $\tilde{\sigma}_y^2$ are therefore identical to those in the time-series specification (with $\tilde{\mu}_z$ identical to $\tilde{\mu}_y$ in the time-series case).

Since the crash frequency and crash size are identical in z_t and $\epsilon_{t,i}$, this implies that the variance decomposition between z_t and $\epsilon_{t,i}$, represented by λ, only captures the non-crash variation. Since the implicit decomposition is 50-50 for the crashes, the total variance (including crashes) attributable to the common factor is somewhat higher than the stated λ of 40%. When λ is estimated in the simulations, the average value is about 0.45.

A3 The ACO block bootstrap

In the simulations in the main paper, we showed that the (skewness-corrected) MLE and bootstrap estimator were only marginally affected by empirically reasonable levels of serial dependence in returns. The results presented in the main text thus suggest that serial correlation is not a major inferential concern. Here we elaborate on this topic and provide some additional simulation results for the block bootstrap estimator proposed by Anarkulova, Cederburg, and O'Doherty (2022, ACO).

ACO extends the FF bootstrap to a block bootstrap method. The idea is identical to the FF bootstrap, but instead of sampling individual returns from $\{y_t\}_{t=1}^n$ (or $\{x_t\}_{t=1}^n$), "blocks" (i.e., contiguous sequences) of returns are instead sampled. This maintains serial dependencies across returns in the bootstrapped samples, provided the blocks are long enough. By maintaining, in the bootstrapped returns, the serial correlation present in the original returns data, the hope is that the block bootstrap estimator will be robust to serial dependence.

ACO provide no formal validation of their block bootstrap procedure and such an analysis is also outside the scope of the current paper. Below, however, we provide simulation results for the ACO block bootstrap and compare the results with those from the estimators proposed in the current study.

We use the same simulation specifications as in the main text: i.i.d. log-normal, i.i.d. log-normal-with-crashes, stochastic volatility (SV), and long-term reversals. The parameterizations are also the same, with the mean and volatility of the gross monthly returns set to $\mu = 1.006$ and $\sigma = 0.06$, respectively. For the long-term reversal specification, monthly log returns follow an MA(q) process. In the main text we set $q = 60$. Here we also consider a specification with $q = 120$. In both cases, the MA(q) process is specified to have a variance ratio equal to 0.8. The details of these specifications are found in the main text and in Online Appendix A2. We thus evaluate the block bootstrap estimator under specifications both with and without serial dependence.

The block bootstrap procedure samples blocks of length l, with replacement. In their main implementation, ACO use randomly distributed block lengths (following a geometric

distribution) with a mean of 120 months. We follow this choice, but as a robustness check we also consider fixed block lengths of 60 months.[3]

Figures A6 and A7 show simulation results in the same format as Figures 1 and 2 in the main text, but with results for the block bootstrap estimators added in as well. The block bootstrap estimators use either randomly distributed block lengths with a mean of 120 months or fixed block lengths of 60 months. As a comparison to the block bootstrap estimators, we show results for the skewness-corrected MLE. In order to keep the graphs uncluttered, we omit results for the MLE and the FF bootstrap estimator. The skewness-corrected MLE is very similar to the FF bootstrap, as seen in the main text, and improves upon the standard MLE without any apparent loss of efficiency. As previously, the simulations are based on samples with $1,440$ monthly observations and $10,000$ simulated samples. Investment horizons of $T = 120$ and $T = 360$ are used.

Following the format of the figures in the main text, Figures A6 and A7 show the estimated quantiles of the long-run return distributions. The solid line shows the true (population) quantiles in each graph. The dash-dotted lines show the median and the 5^{th} and 95^{th} percentiles of the block bootstrap estimates of each quantile, using a random block length with a mean of 120 months. The dotted lines show the corresponding results for the block bootstrap estimator with a fixed block length of 60 months. The dashed line shows the median estimates for the skewness-corrected MLE and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the skewness-corrected ML estimates. Additional results are shown in Tables A2 and A3, which follow the same format as Tables 1 and 2 in the main text.

Panels A and B in Figures A6 and A7 show results for the i.i.d. log-normal and i.i.d. log-normal-with-crashes distributions. For both of these distributions, the skewness-corrected MLE is essentially median unbiased. The block bootstrap estimator exhibits a small but noticeable bias, especially when using the longer random block lengths. The differences

[3]In non-reported results, we also considered fixed block lengths of 120 months and random block lengths with a mean of 60 months. The results for the fixed length of 60 months and the random lengths with a mean of 120 months lead to the most disparate results. The other two alternatives thus end up somewhere in between. We report results for only two different block lengths to keep the presentation manageable.

between the skewness-corrected MLE and the block bootstrap estimators are easily seen in the lines capturing the 95^{th} percentiles of the estimates. These are quite clearly different for the block bootstrap estimators and the skewness-corrected MLE (represented by the right-hand edge of the shaded region). There is a clear tendency for the block bootstrap estimators to over-estimate the lower quantiles (i.e., estimate them with an upward bias). In the case of i.i.d. returns, using the block bootstrap therefore comes at a cost, albeit a small one.

Tables A2 and A3 further support this interpretation. The (median) biases for the block bootstrap estimators are larger than for the skewness-corrected MLE and so are the median absolute errors. Interestingly, these results are very similar across the log-normal and log-normal-with-crashes specifications. The bias in the block bootstrap is therefore not driven by non-normality. The results for the SV specification, shown in Panel C of the figures and tables, are very similar to those for the two i.i.d. processes.

Panels D and E show results for the two MA specifications. Given that these specifications exhibit serial correlation, one would expect the block bootstrap estimators to provide less biased estimates, at the possible cost of greater dispersion. This is also what we find. The bias is smaller for the block bootstrap estimator but, as seen in the tables, the skewness-corrected MLE still tends to dominate in terms of median absolute error. There is also still a tendency to overestimate the lower quantiles when using the block bootstrap estimators, which is especially clear for the 30-year horizon shown in Panels D and E of Figure A7. This bias is more severe for the longer (random) block length.

In Tables A2 and A3, we also show the coverage rates of confidence intervals centered on the block bootstrap estimates (see top rows in each panel of the tables). That is, we calculate the same confidence bands as used for the other estimators, but center them instead on the block bootstrap estimators. There is no formal result supporting such a confidence band, and poor coverage rates do not in and of themselves invalidate the block bootstrap estimator. However, since the median block bootstrap estimates are still quite similar to the median skewness-corrected ML estimates, differences in the coverage rates still give some indication of differing dispersion in the estimates, as already evidenced

in the figures. As seen from the tables, the coverage rates for confidence intervals based on the block bootstrap estimates tend to be much worse than those centered on the skewness-corrected ML estimates. This is especially true for $T = 120$.

To sum up, there is little evidence that the block bootstrap estimators perform significantly better than the skewness-corrected MLE in the presence of serial correlation. The block bootstrap can reduce the bias somewhat, but at the expense of greater dispersion. When returns are i.i.d., or a martingale difference sequence in the form of a stochastic volatility process, the block bootstrap is clearly dominated by the skewness-corrected MLE. In all cases, the latter tends to dominate in median absolute error terms.

A4 Additional Empirical results

In this section, we add further information and results for the empirical analysis.

Figure A8 provides detailed information on the full set of countries included in the DMS data set and for which time periods the data are available for a given country.

Tables A4-A9 show tabulated results for the country-specific estimates for 10-year and 30-year horizons, using the three different estimators for calculating the point estimates of the long-run distributions. Tables A4 and A5 use the MLE. Tables A6 and A7 use the skewness-corrected MLE. Finally, Tables A8 and A9 show results for the FF bootstrap estimator. In each case, 90% confidence intervals are presented along with the point estimates. Global-versus-individual-country comparison figures, analogous to Figure 4 in the main text, are presented for each country in the original DMS data set in Figures A9 and A10.

Table A10 replicates the analysis in Table 4 in the main text, using the unbalanced panel of all 32 countries that the DMS data set currently includes (see Figure A8 for a full list of all countries).

Finally, Table A11 presents the ACO block bootstrap results from estimating the global long-run returns, using the balanced 21-country panel. As a comparison, the FF bootstrap results are also shown. The results confirm that differences between the estimators tend to be quite small, especially when compared to the uncertainty in the estimates. However, the

tendency for the block bootstrap estimator to produce estimates of the lower quantiles that are larger than the corresponding estimates from the FF bootstrap is evident. Compared to the sampling uncertainty, all differences are small and well within the confidence intervals.

A5 Formal results for estimation and testing with panel data

In this section, we provide formal results and additional details for our discussion of panel data in Section 4 in the main text. All proofs for the results in this section are gathered in Section A5.3 below. The analysis is based on the skewness-corrected MLE. If one assumes zero skewness in the one-period distributions, all results also apply to the plain MLE without skewness correction.

A5.1 Asymptotic distribution of panel estimator

This section provides the theoretical foundation for the panel-based confidence intervals defined in Section 4.1 in the main text.

We use the notation introduced in Section 4.1 in the main text and operate under the assumptions specified there. In particular, we suppose that we observe i.i.d. observations of the random vector $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{K,t})$, where all components share the same marginal distribution with mean μ_y, variance σ_y^2, and skewness γ_y. We do not restrict the cross-sectional dependence of the components of $\mathbf{y}_t$.

Proposition A1. *Suppose that the vector $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{K,t})$ is i.i.d. and that all its components share the same marginal distribution, which has a continuous density, a positive variance, and eight moments. Then, for any $\tau \in (0,1)$,*

$$\begin{aligned}
&\left(\widehat{Q}_{Y_T}^{ML\text{-}skew,pool}(\tau) - Q_{Y_T}(\tau)\right)/\sigma_y \\
=&\frac{T}{nK}\sum_{t=1}^{n}\sum_{i=1}^{K}\left\{\frac{y_{i,t}-\mu_y}{\sigma_y} + \frac{1}{2\sqrt{T}}\left(\left(\frac{y_{i,t}-\mu_y}{\sigma_y}\right)^2 - 1\right)\Phi^{-1}(\tau)\right\} \qquad \text{(A26)} \\
&+ O_p\left(n^{-1/2} + T^{-1/2}\right).
\end{aligned}$$

This expansion immediately implies the following distributional result:

Proposition A2. *Suppose that the vector* $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{K,t})$ *is i.i.d. and that all its components share the same marginal distribution, which has a continuous density, a positive variance, and eight moments. Let*

$$V_{t,\tau} = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{y_{i,t}-\mu_y}{\sigma_y} + \frac{1}{2\sqrt{T}}\left(\left(\frac{y_{i,t}-\mu_y}{\sigma_y}\right)^2 - 1\right)\Phi^{-1}(\tau)\right) \tag{A27}$$

and $\sigma^2_{V,\tau} = \mathbb{E}\left[V^2_{t,\tau}\right] = \operatorname{var}(V_{t,\tau})$. *If the sequence* $T = T(n)$ *diverges fast enough such that* $T^3/n \to \infty$ *then, for any* $\tau \in (0,1)$, *we have*

$$\frac{\sqrt{n}}{T\sigma_y\sigma_{V,\tau}}\left(\widehat{Q}^{\textit{ML-skew,pool}}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right) \Rightarrow N(0,1). \tag{A28}$$

To use this result for practical inference, we need a consistent estimator of $\sigma^2_{V,\tau}$. To see why such an estimator is furnished by $\hat{\sigma}^2_{V,\tau}$, defined in Section 4.1 in the main text, note that $\widehat{V}_{t,\tau}$ is an uncentered sample counterpart of $V_{t,\tau}$, which replaces the population quantities μ_y and σ_y by consistent estimators. Under the assumptions of Proposition A2,

$$\hat{\sigma}^2_{V,\tau} = \sigma^2_{V,\tau} + O_p\left(n^{-1/2}\right). \tag{A29}$$

Thus, by standard arguments and the result of Proposition A1, we have

$$\frac{\sqrt{n}}{T\hat{\sigma}^{\text{pool}}_y\hat{\sigma}_{V,\tau}}\left(\widehat{Q}^{\text{ML-skew,pool}}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right) \Rightarrow N(0,1). \tag{A30}$$

This result establishes the asymptotic validity of the confidence interval defined in Section 4.1 in the main text.

Example with a single factor We now provide further details for our example with a single factor from the main text. The one-factor model is given by

$$y_{t,i} = \mu_y + z_t + \epsilon_{t,i}, \tag{A31}$$

where z_t is a common factor and $\epsilon_{t,i}$ is an idiosyncratic error term. Plugging the one-factor model into the expression for $U_{t,\tau}$, we obtain

$$\begin{aligned} V_{t,\tau} =& \frac{1}{K}\sum_{i=1}^{K}\left(\frac{y_{i,t}-\mu_y}{\sigma_y}+\frac{1}{2\sqrt{T}}\left(\left(\frac{y_{i,t}-\mu_y}{\sigma_y}\right)^2-1\right)\Phi^{-1}(\tau)\right) \\ =& z_t/\sigma_y+\frac{1}{K}\sum_{i=1}^{K}\epsilon_{i,t}/\sigma_y+\frac{1}{2\sqrt{T}}\left(\left(z_t/\sigma_y\right)^2-1\right)\Phi^{-1}(\tau) \\ &+\frac{\Phi^{-1}(\tau)}{2\sqrt{T}K}\sum_{i=1}^{K}\left(\epsilon_{i,t}/\sigma_y\right)^2+\left(z_t/\sigma_y\right)\frac{\Phi^{-1}(\tau)}{\sqrt{T}K}\sum_{i=1}^{K}\epsilon_{i,t}/\sigma_y. \end{aligned} \tag{A32}$$

It is now straightforward to derive that

$$\begin{aligned} \text{var}(V_{t,\tau}) =& \mathbb{E}\left[z_t/\sigma_y\right]^2+\frac{1}{K}\sum_{i=1}^{K}\mathbb{E}\left[\epsilon_{i,t}/\sigma_y\right]^2+O\left(T^{-1/2}\right) \\ =& \text{var}\left(z_t\right)/\sigma_y^2+\frac{\text{var}\left(\epsilon_{i,t}\right)/\sigma_y^2}{K}+O\left(T^{-1/2}\right) \\ =& \lambda+\frac{1-\lambda}{K}+O\left(T^{-1/2}\right). \end{aligned} \tag{A33}$$

A5.2 Testing for differences in the distributions of long-run returns

We observe K countries, indexed by $i=1,\ldots,K$, over n time periods. The log return of country i at time t is denoted $y_{i,t}$. We assume that the countries $k=2,\ldots,K$ share the same marginal return distribution. The τ-quantile of the T-period log return to investing in a one-unit asset in one of the countries $k=2,\ldots,K$ is denoted $Q_{Y_T,-1}(\tau)$. For all $i=2,\ldots,K$, the distribution of the one-period log returns has the moments

$$\mathbb{E}\left[y_{i,t}\right]=\mu_{y,-1},\ \text{var}(y_{i,t})=\sigma_{y,-1}^2 \quad \text{and} \quad \mathbb{E}\left[\frac{y_{i,t}-\mu_{y,-1}}{\sigma_{y,-1}}\right]^2=\gamma_{y,-1}. \tag{A34}$$

Country 1 faces a potentially different return distribution and its T-period log returns have quantile function $Q_{Y_T,1}(\tau)$. Country 1's one-period log returns have mean $\mu_{y,1}$, variance $\sigma_{y,1}^2$ and skewness $\gamma_{y,1}$. We assume that the return vector $\mathbf{y}_t=(y_{t,1},y_{t,2},\ldots,y_{t,K})$ is identically and independently distributed over time, but we allow for any kind of cross-sectional dependence. In particular, the variance matrix of $\mathbf{y}_t$ may not be a diagonal matrix.

Proposition A3. *Assume that the vector* $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{K,t})$ *is i.i.d. and that the components* $y_{2,t}, \ldots, y_{K,t}$ *have the same marginal distribution. Suppose that the* $y_{i,t}$ *have continuous densities and eight moments and let* $\tau = (0, 1)$ *denote a quantile of interest. Let* $\widehat{Q}^{ML\text{-}skew}_{Y_T,1}(\tau)$ *denote the skewness-corrected ML estimator based on the return series of country 1 and let* $\widehat{Q}^{ML\text{-}skew,pool}_{Y_T,-1}(\tau)$ *denote the leave-one-out pooled skewness-corrected ML estimator that excludes country 1. If the sequence* $T = T(n)$ *diverges fast enough such that* $T^3/n \to \infty$ *and*

$$\operatorname{var}\left(y_{1,t} - \frac{1}{K-1}\sum_{i=2}^{K} y_{i,t}\right) > 0 \tag{A35}$$

then

$$\frac{\sqrt{n}}{T\sigma_{W,\tau}}\left(\widehat{Q}^{ML\text{-}skew}_{Y_T,1}(\tau) - \widehat{Q}^{ML\text{-}skew,pool}_{Y_T,-1}(\tau) - \left(Q_{Y_T,1}(\tau) - Q_{Y_T,-1}(\tau)\right)\right) \Rightarrow N(0,1), \tag{A36}$$

where $\sigma^2_{W,\tau} = \operatorname{var}(W_{t,\tau})$ *and*

$$\begin{aligned} W_{t,\tau} =& y_{1,t} - \mu_{y,1} - \frac{1}{K-1}\sum_{i=2}^{K}(y_{i,t} - \mu_{y,-1}) \\ &+ \frac{\Phi^{-1}(\tau)}{2\sqrt{T}}\left\{\sigma_{y,1}\left(\left(\frac{y_{1,t}-\mu_{y,-1}}{\sigma_{y,1}}\right)^2 - 1\right) - \frac{\sigma_{y,-1}}{(K-1)}\sum_{i=2}^{K}\left(\left(\frac{y_{i,t}-\mu_{y,-1}}{\sigma_{y,-1}}\right)^2 - 1\right)\right\}. \end{aligned} \tag{A37}$$

To define an estimator of $\sigma^2_{W,\tau}$, let $\hat{\mu}_{y,1}$ and $\hat{\sigma}^2_{y,1}$ denote the sample mean and sample variance of the return series of country 1:

$$\hat{\mu}_{y,1} = \frac{1}{n}\sum_{t=1}^{n} y_{1,t} \quad \text{and} \quad \hat{\sigma}^2_{y,1} = \frac{1}{n}\sum_{t=1}^{n}(y_{1,t} - \hat{\mu}_{y,1})^2. \tag{A38}$$

Let $\hat{\mu}_{y,-1}$ and $\hat{\sigma}^2_{y,-1}$ denote the leave-one-out pooled sample mean and variance that exclude country 1:

$$\hat{\mu}_{y,-1} = \frac{1}{n(K-1)}\sum_{t=1}^{n}\sum_{i=2}^{K} y_{i,t} \quad \text{and} \quad \hat{\sigma}^2_{y,-1} = \frac{1}{n(K-1)}\sum_{t=1}^{n}\sum_{i=2}^{K}(y_{i,t} - \hat{\mu}_{y,-1})^2. \tag{A39}$$

And let

$$\begin{aligned}\widehat{W}_{t,\tau} =& y_{1,t} - \frac{1}{K-1}\sum_{i=2}^{K} y_{i,t} \\ &+ \frac{\Phi^{-1}(\tau)}{2\sqrt{T}}\left\{\hat{\sigma}_{y,1}\left(\frac{y_{1,t}-\hat{\mu}_{y,-1}}{\hat{\sigma}_{y,1}}\right)^2 - \frac{\hat{\sigma}_{y,-1}}{(K-1)}\sum_{i=2}^{K}\left(\frac{y_{i,t}-\hat{\mu}_{y,-1}}{\hat{\sigma}_{y,-1}}\right)^2\right\},\end{aligned} \tag{A40}$$

denote an uncentered sample counterpart of $W_{t,\tau}$. Now, an estimator of $\sigma^2_{W,\tau}$ is given by

$$\hat{\sigma}^2_{W,\tau} = \frac{1}{n}\sum_{t=1}^{n}\left(\widehat{W}_{t,\tau} - \frac{1}{n}\sum_{s=1}^{n}\widehat{W}_{s,\tau}\right)^2. \tag{A41}$$

This estimator is consistent under the assumptions of Proposition A3. Therefore, by Proposition A3, a test with test statistic

$$T_\Delta(\tau) = \frac{\sqrt{n}}{T\hat{\sigma}_{W,\tau}}\left(\widehat{Q}^{\text{ML-skew}}_{Y_T,1}(\tau) - \widehat{Q}^{\text{ML-skew,pool}}_{Y_T,-1}(\tau) - \Delta(\tau)\right) \Rightarrow N(0,1) \tag{A42}$$

that rejects the null hypothesis $H_0 : Q_{Y_T,1}(\tau) - Q_{Y_T,-1}(\tau) = \Delta(\tau)$ if $|T(\tau)| > \Phi^{-1}(1-\alpha/2)$ has asymptotic size α.

Example with a single factor The power of the $T_\Delta(\tau)$-test is governed by the magnitude of $T\hat{\sigma}_{W,\tau}/\sqrt{n}$. By contrast, the width of the confidence interval for the long-run (log) return distribution of a single-country is determined by $T\hat{\sigma}_y\psi_T(\tau)/\sqrt{n}$. Thus, whether the objective is to learn about the return distribution or about *differences* in return distributions, the precision of inference is constrained by the same asymptotic rate, $T/\sqrt{n}$. However, the terms $\hat{\sigma}_y\psi_T(\tau) \approx \sigma_y$ (since $\psi_T(\tau) \approx 1$ for large T) and $\hat{\sigma}_{W,\tau} \approx \sigma_{W,\tau}$, may differ substantially. We illustrate this based on the one-factor model,

$$\begin{aligned} y_{1,t} =& \mu_{y,1} + z_t + \epsilon_{1,t} \quad \text{and} \\ y_{i,t} =& \mu_{y,-1} + z_t + \epsilon_{i,t} \quad \text{for } i = 2,\ldots,K, \end{aligned} \tag{A43}$$

where the factor z_t is a common factor for all countries and $\mathbb{E}[z_t] = \mathbb{E}[\epsilon_{i,t}] = 0$ and the $\epsilon_{i,t}$ are independent across countries and identically distributed among countries $2,\ldots,K$.

Plugging the factor model into the expression for $W_{t,\tau}$, we obtain

$$\begin{aligned} W_{t,\tau} =& \epsilon_{1,t} - \frac{1}{K-1}\sum_{i=2}^{K}\epsilon_{i,t} \\ &+ \frac{\sigma_{y,1}\Phi^{-1}(\tau)}{2\sqrt{T}}\left(\left((z_t+\epsilon_{1,t})/\sigma_{y,1}\right)^2 - 1\right) \\ &- \frac{\sigma_{y,-1}\Phi^{-1}(\tau)}{2\sqrt{T}}\sum_{i=2}^{K}\left(\left((z_t+\epsilon_{i,t})/\sigma_{y,-1}\right)^2 - 1\right). \end{aligned} \tag{A44}$$

Therefore,

$$\begin{aligned} \sigma^2_{W,\tau} =& \operatorname{var}\left(W_{t,\tau}\right) \\ =& \operatorname{var}\left(\epsilon_{1,t}\right) + \frac{\operatorname{var}\left(\epsilon_{2,t}\right)}{K-1} + O\left(T^{-1/2}\right) \\ =& \sigma^2_{y,1}(1-\lambda_1) + \frac{\sigma^2_{y,-1}(1-\lambda_{-1})}{K-1}, \end{aligned} \tag{A45}$$

where $\lambda_1 = \operatorname{var}(z_t)/\sigma^2_{y,1}$ and $\lambda_2 = \operatorname{var}(z_t)/\sigma^2_{y,-1}$ are the variance shares of the common factor in the two return distributions that we are comparing. Under the null hypothesis that all countries face the same return distribution, $\sigma^2_{y,1} = \sigma^2_{y,-1} = \sigma^2_y$ and $\lambda_1 = \lambda_2 = \lambda$. It follows that

$$\hat{\sigma}^2_{W,\tau} \approx \sigma^2_{W,\tau} = \sigma^2_y(1-\lambda)\frac{K}{(K-1)} \quad \text{and} \quad \hat{\sigma}^2_{y,1}\psi^2_T(\tau) \approx \sigma^2_y. \tag{A46}$$

For example, if $\lambda = 0.3$ and $K \geq 4$, then $\sigma^2_{W,\tau} < \sigma^2_y$ and inference on the distribution of country 1 is less precise than inference on the difference in distributions for panels with at least 4 countries. In the full-sample empirical analysis in the main text, $\lambda \approx 0.4$ and $K = 21$. In this case, $\sigma^2_{W,\tau} \approx \sigma^2_y \times (1-0.4) \times (21/20) = 0.42\sigma^2_y$. It is therefore not surprising if the tests of equality appear "more precise" than what one might expect from the very wide confidence intervals around the individual countries.

A5.3 Proofs for panel results

Proof of Proposition A1. We start by deriving an expansion for the pooled variance estimator. First, note that Markov's inequality implies that

$$\left(\hat{\mu}_y^{\text{pool}} - \mu_y\right)/\sigma_y = \frac{1}{nK}\sum_{t=1}^{n}\sum_{i=1}^{K}\frac{y_{i,t}-\mu_y}{\sigma_y} = O_p\left(n^{-1/2}\right).$$

Therefore, we have that

$$\begin{aligned}
\hat{\sigma}_y^{2,\text{pool}}/\sigma_y^2 =& \frac{1}{nK}\sum_{t=1}^{n}\sum_{i=1}^{K}\left(\frac{y_{i,t}-\hat{\mu}_y^{\text{pool}}}{\sigma_y}\right)^2 \\
=& \frac{1}{nK}\sum_{t=1}^{n}\sum_{i=1}^{K}\left(\frac{y_{i,t}-\mu_y}{\sigma_y}\right)^2 - \left(\frac{\hat{\mu}_y^{\text{pool}}-\mu_y}{\sigma_y}\right)^2 \\
=& \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{1}{K}\sum_{i=1}^{K}\left(\frac{y_{i,t}-\mu_y}{\sigma_y}\right)^2\right\} + O_p\left(n^{-1}\right).
\end{aligned}$$

In particular, by Markov's inequality,

$$\hat{\sigma}_y^{2,\text{pool}}/\sigma_y^2 = 1 + O_p\left(n^{-1/2}\right).$$

Noting that

$$\sqrt{w} - 1 = \frac{1}{2}(w-1) + O((w-1)^2),$$

for $w > 0$, we obtain the stochastic expansion

$$\begin{aligned}
\hat{\sigma}_y^{\text{pool}}/\sigma_y - 1 =& \frac{1}{2}\left(\frac{\hat{\sigma}_y^{\text{pool},2}}{\sigma_y^2} - 1\right) + O_p\left(n^{-1}\right) \\
=& \frac{1}{2}\left(\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{1}{K}\sum_{i=1}^{K}\left(\frac{y_{i,t}-\mu_y}{\sigma_y}\right)^2\right\} - 1\right) + O_p\left(n^{-1}\right).
\end{aligned}$$

By standard arguments, we obtain $\hat{\gamma}_y^{\text{pool}} - \gamma_y = O_p(n^{-1/2})$. By Lemma A4, we can therefore write

$$\begin{aligned}
&\frac{\sqrt{n}}{T\sigma_y}\left(\widehat{Q}_{Y_T}^{\text{ML-skew,pool}}(\tau) - Q_{Y_T}(\tau)\right) \\
=&\sqrt{n}\left(\hat{\mu}_y^{\text{pool}} - \mu_y\right)/\sigma_y + \sqrt{n}/\sqrt{T}\left(\left(\hat{\sigma}_y^{\text{pool}}/\sigma - 1\right)\right)\Phi^{-1}(\tau) + O_p(T^{-1} + \left(T^3/n\right)^{-1/2}) \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_{t,\tau} + O_p\left(T^{-1} + \left(T^3/n\right)^{-1/2}\right).
\end{aligned}$$

□

Proof of Proposition A2. The result follows from Proposition A1 and the Lindeberg central limit theorem. □

Proof of Proposition A3. By standard arguments we obtain

$$\hat{\sigma}_{y,1} = \sigma_{y,1} + O_p\left(n^{-1/2}\right) \quad \text{and} \quad \hat{\gamma}_{y,1} = \gamma_{y,1} + O_p\left(n^{-1/2}\right).$$

Hence,

$$\begin{aligned}
&\widehat{Q}_{Y_T,1}^{\text{ML-skew}}(\tau) - Q_{Y_T,1}(\tau) \\
=&T\left(\hat{\mu}_{y,1} - \mu_{y,1}\right) + \sqrt{T}\left(\hat{\sigma}_{y,1} - \sigma_{y,1}\right)\Phi^{-1}(\tau) + O_p\left(n^{-1/2} + T^{-1/2}\right).
\end{aligned}$$

Therefore, by Lemma A2,

$$\begin{aligned}
&\frac{\sqrt{n}}{T}\left(\widehat{Q}_{Y_T,1}^{\text{ML-skew}}(\tau) - Q_{Y_T,1}(\tau)\right) \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{y_{1,t} - \mu_{y,1} + \frac{\sigma_{y,1}}{2\sqrt{T}}\left(\left(\frac{y_{1,t} - \mu_{y,1}}{\sigma_{y,1}}\right)^2 - 1\right)\Phi^{-1}(\tau)\right\} \\
&+ O_p\left(T^{-1/2} + (T^3/n)^{-1/2}\right).
\end{aligned}$$

By Proposition A1,

$$\begin{aligned}
&\frac{\sqrt{n}}{T}\left(\widehat{Q}_{Y_T}^{\text{ML-skew,pool}}(\tau) - Q_{Y_T}(\tau)\right) \\
=&\frac{1}{(K-1)\sqrt{n}}\sum_{t=1}^{n}\sum_{i=2}^{K}\left\{y_{i,t} - \mu_y + \frac{\sigma_y}{2\sqrt{T}}\left(\left(\frac{y_{i,t}-\mu_y}{\sigma_y}\right)^2 - 1\right)\Phi^{-1}(\tau)\right\} \\
&+ O_p\left(T^{-1} + (T^3/n)^{-1/2}\right).
\end{aligned}$$

The conclusion follows by subtracting the two expansions and applying the Lindeberg central limit theorem. □

A6 Consistency for gross returns

Proposition A1 (**Consistent estimation in normal model**). *Suppose that single-period log returns, y_t, are i.i.d. normal or, equivalently, that single-period gross-returns, x_t, are i.i.d. log-normal. Then the following results hold:*

1. *The ML quantile estimator $\widehat{Q}_{Y_T}^{ML}(\tau)$ is consistent for $Q_{Y_T}(\tau)$ if and only if $T(n)/\sqrt{n} \to 0$ as $n \to \infty$.*

2. *The ML quantile estimator $\widehat{Q}_{X_T}^{ML}(\tau)$ is consistent for $Q_{X_T}^{ML}(\tau)$ if and only if*

$$\limsup_{n\to\infty} \frac{2T(n)\mu_y}{\log n} < 1 \quad \text{as } n \to \infty. \tag{A47}$$

The first part of Proposition A1 simply restates the result from Proposition 1 in the main text. The second part of Proposition A1 states a rate condition for consistent estimation of the gross long-run return. It is much stronger than the corresponding condition for the log return, requiring T to be small compared to $\log n$ (instead of just $\sqrt{n}$). To understand why a stronger condition is needed, note that, under the log-normal model, the variance of a single T-period gross return is equal to

$$\text{var}\,(X_T) = (1 - \exp(-T\sigma_y^2))\exp(2T(\mu_y + \sigma_y^2)), \tag{A48}$$

and hence grows exponentially with T, instead of linearly as in the log return case. This makes "direct" estimation of the long-run gross return much more challenging than the corresponding estimation of the log return. This is then also reflected in the rate condition for the MLE.

A7 Technical derivations

A7.1 Proofs of main results

Proof of Proposition A1 (and thus Proposition 1). We first prove the first claim in the proposition. Let

$$\psi_T^2(\tau) = 1 + \frac{1}{2T}\left(\Phi^{-1}(\tau)\right)^2.$$

Define the positive sequence

$$a_n = \frac{T\sigma_y\psi_T(\tau)}{\sqrt{n}}.$$

By Lemma A2,

$$a_n^{-1}\left(\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right) \Rightarrow N(0,1).$$

If $a_n \to 0$ then this implies immediately

$$\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau) = o_p(1).$$

On the other hand, if $a_n \to 0$ does not hold, then there is positive constant c such that $a_n \geq c$ along a subsequence. Pass to this subsequence. Along the subsequence we have

$$\left|\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right| \geq ca_n^{-1}\left(\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right)$$

and therefore, for any $\eta > 0$,

$$P\left(\left|\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right| > \eta\right) \geq P\left(\left|a_n^{-1}\left(\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right)\right| > \eta/c\right) \to 2\Phi\left(-\eta/c\right),$$

a positive number. Here, the convergence follows since the subsequence inherits convergence to a normal limit from the original sequence. Noting that $a_n \to 0$ if and only if $T/\sqrt{n} \to 0$ concludes the proof of the first claim.

Turning to the second claim, suppose first that $\limsup_{n\to\infty} \frac{2T\mu_y}{\log n} < 1$. If $T = T(n)$ is bounded then consistency follows trivially. Suppose therefore that T is not bounded and pass to a subsequence along which T diverges. We can write

$$\widehat{Q}_{X_T}(\tau) - Q_{X_T}(\tau) = \exp\left(Q_{Y_T}(\tau)\right)\left\{\exp\left(\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right) - 1\right\}.$$

$\limsup_{n\to\infty} \frac{2T\mu_y}{\log n} < 1$ implies that along all subsequences (in particular the one we just passed to) $T/\sqrt{n} \to 0$. Under this condition, the first claim of the proposition implies

$$\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau) = o_p(1).$$

In particular, $\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau)$ is bounded by $\log 1$ with probability approaching one. Hence, setting $x = \widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau)$ in the inequality $|e^x - 1| \leq |x|e^{|x|}$ implies

$$\left|\exp(Q_{Y_T}(\tau))\left[\left(\widehat{Q}_{X_T}(\tau) - Q_{X_T}(\tau)\right) - 1\right]\right| \leq \exp(Q_{Y_T}(\tau))\left|\widehat{Q}_{X_T}(\tau) - Q_{X_T}(\tau)\right|$$

Arguing similarly as above, we have

$$\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau) = O_p\left(T/\sqrt{n}\right).$$

Therefore, there is a constant $C > 0$ such that with probability approaching one,

$$\begin{aligned}
&\exp(Q_{Y_T}(\tau)) \left| \widehat{Q}_{X_T}(\tau) - Q_{X_T}(\tau) \right| \\
\leq & C(T/\sqrt{n}) \exp\left(T\mu_y + \sqrt{T}\sigma_y \Phi^{-1}(\tau) \right) \\
\leq & C(T/\sqrt{n}) \exp\left(T\mu_y \left(1 + C/\sqrt{T} \right) \right) \\
\leq & C \exp\left(T\mu_y \left(1 + C/\sqrt{T} + (\log T)/(T\mu_y) \right) - \frac{1}{2}\log n \right) \\
= & C \exp\left\{ -\frac{1}{2}\log n \left(1 - \frac{2T\mu_y}{\log n} \left(1 + C/\sqrt{T} + (\log T)/(T\mu_y) \right) \right) \right\}.
\end{aligned}$$

The condition, $\limsup_{n\to\infty} \frac{2T\mu_y}{\log n} < 1$ guarantees that

$$1 - \frac{2T\mu_y}{\log n} \left(1 + C/\sqrt{T} + (\log T)/(T\mu_y) \right) > 0$$

for n large enough. Hence, combining the inequalities above, there is $c > 0$ such that with probability approaching one

$$\left| \widehat{Q}_{X_T}(\tau) - Q_{X_T}(\tau) \right| \leq C n^{-c}.$$

This concludes the proof of sufficiency. To prove necessity, suppose that $\limsup_{n\to\infty} \frac{2T\mu_y}{\log n} \geq 1$. The inequality $e^x - 1 \geq x$ implies

$$\begin{aligned}
\widehat{Q}_{X_T}(\tau) - Q_{X_T}(\tau) = & \exp\left(Q_{Y_T}(\tau) \right) \left\{ \exp\left(\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau) \right) - 1 \right\} \\
& \geq \exp\left(T\mu_y + \sqrt{T}\sigma_y \Phi^{-1}(\tau) \right) \left(\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau) \right).
\end{aligned}$$

Arguing as above, we have

$$\frac{\sqrt{n}}{T\sigma_y} \left(\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau) \right) \Rightarrow N(0,1)$$

and hence for $\eta > 0$

$$P\left(\frac{\sqrt{n}}{T\sigma_y} \left(\widehat{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau) \right) > \eta \right) \to \Phi(-\eta)$$

Therefore, with probability approaching $\Phi(-\eta)$,

$$\begin{aligned}&\widehat{Q}_{X_T}(\tau) - Q_{X_T}(\tau)\\ &\geq \exp\left\{\frac{1}{2}\log n\left[\frac{2T\mu_y}{\log n}\left(1+\sigma_y\Phi^{-1}(\tau)/(\mu_y\sqrt{T})+\log T/(T\mu_y)\right)-1\right]\right\}\eta\end{aligned}$$

and $\limsup_{n\to\infty}\frac{2T\mu_y}{\log n}\geq 1$ ensures that $\widehat{Q}_{X_T}(\tau)-Q_{X_T}(\tau)$ is bounded away from zero along a subsequence with probability approaching $\Phi(-\eta)>0$. This proves necessity and concludes the proof of the second claim. □

Proof of Proposition 2. This follows immediately from Lemma A2. □

Proof of Proposition 3. By Lemma A1, it suffices to prove the proposition for the bootstrap estimator.

Any sequence $T(n)$ is either bounded, diverges to infinity, or can be divided into a bounded subsequence and a diverging subsequence. To prove the proposition, it suffices therefore to prove the claim for arbitrary bounded subsequences and arbitrary diverging subsequences.

Bounded subsequence: We first consider the case of a bounded subsequence and pass to this subsequence, i.e., we assume $T(n)\leq\bar{T}$. Necessity holds trivially since $T/\sqrt{n}\leq\bar{T}/\sqrt{n}\to 0$. Therefore, we only have to prove sufficiency. Let f_{Y_T} denote the density of Y_T. The assumption that y_t has a density and is supported on an interval, implies that, for fixed T, the density of its T-fold convolution is supported on an interval. Therefore, there is $\epsilon>0$ such that, for all y with $|y-Q_{Y_T}(\tau)|\leq\epsilon$, we have $f_{Y_T}(y)\geq\epsilon$. Since $T\leq\bar{T}$ we can find $\epsilon>0$ such that this property holds for all T along the $T(n)$ sequence. Let F_{Y_T} denote the distribution function of Y_T and let $\hat{F}_{Y_T^*}(a)=P(Y^*\leq a\mid\mathcal{Y}_n)$. Let

$$\delta_n=(2C_1/\epsilon)\frac{T}{\sqrt{n}}\sqrt{\log\frac{n}{T^2}}.$$

From here on, assume that n is large enough such that $\delta_n<\epsilon$. Define the event

$$\mathcal{E}_n=\left\{\sup_{a\in\mathbb{R}}\left|F\left(a\right)-\hat{F}_{Y_T^*}\left(a\right)\right|\leq\delta_n(\epsilon/2)\right\}.$$

Suppose that $\widehat{Q}^{\text{boot}}_{Y_T}(\tau) \leq Q_{Y_T}(\tau) - \delta_n$. Then

$$F\left(\widehat{Q}^{\text{boot}}_{Y_T}\right) - \tau \leq F_{Y_T}(Q_{Y_T}(\tau) - \delta_n) - \tau = -\int_{Q_{Y_T}(\tau)-\delta_n}^{Q_{Y_T}(\tau)} f_{Y_T}(x)\,dx \leq -\delta_n\epsilon.$$

Since $\hat{F}_{Y_T^*}(\widehat{Q}^{\text{boot}}_{Y_T}(\tau)) = \tau$,

$$F\left(\widehat{Q}^{\text{boot}}_{Y_T}\right) - \hat{F}_{Y_T^*}\left(\widehat{Q}^{\text{boot}}_{Y_T}\right) \leq -\delta_n\epsilon.$$

This cannot happen on $\mathcal{E}_n$.

Conversely, suppose that $\widehat{Q}^{\text{boot}}_{Y_T}(\tau) \geq Q_{Y_T}(\tau) + \delta_n$. Then

$$F\left(\widehat{Q}^{\text{boot}}_{Y_T}\right) - \tau \geq F_{Y_T}(Q_{Y_T}(\tau) + \delta_n) - \tau = \int_{Q_{Y_T}(\tau)}^{Q_{Y_T}(\tau)+\tau} f_{Y_T}(x)\,dx \geq \delta_n\epsilon.$$

Therefore,

$$F\left(\widehat{Q}^{\text{boot}}_{Y_T}\right) - \hat{F}_{Y_T^*}\left(\widehat{Q}^{\text{boot}}_{Y_T}\right) \geq \delta_n\epsilon.$$

This, too, cannot happen on $\mathcal{E}_n$. In summary,

$$P\left(\left|\widehat{Q}^{\text{boot}}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right| \geq \delta_n\right) \leq P\left(\mathcal{E}_n^{\mathsf{c}}\right) \to 0,$$

where convergence follows from Lemma A3.

Diverging subsequence: We now turn to the case of a diverging subsequence and pass to this subsequence, i.e., we assume $T(n) \to \infty$. Let

$$\begin{aligned} W_n =& T\hat{\mu}_y + \sqrt{T}\hat{\sigma}_y\Phi^{-1}(\tau) - \left(T\mu_y + \sqrt{T}\sigma_y\Phi^{-1}(\tau)\right), \\ a_n =& \left(T\nu_T(\tau)\sigma_y\right)/\sqrt{n}, \end{aligned}$$

where $\nu_T(\tau)$ is defined in Lemma A2. By Lemma A5 we have

$$\widehat{Q}^{\text{boot}}_{Y_T}(\tau) - Q_{Y_T}(\tau) = W_n + O_p\left(n^{-1/2} + T^{-1/2}\right) = o_p(1).$$

Therefore, we have $\widehat{Q}^{\text{boot}}_{Y_T}(\tau) - Q_{Y_T}(\tau) \xrightarrow{p} 0$ if and only if $W_n \xrightarrow{p} 0$. By Lemma A2, $a_n^{-1} W_n \to N(0,1)$. Arguing similarly as in the proof of Proposition 1, it can be shown that W_n vanishes in probability if and only if $a_n \to 0$ or, equivalently, if and only if $T/\sqrt{n} \to 0$. □

Proof of Proposition 4. Let

$$\begin{aligned} W_n =& T\hat{\mu}_y + \sqrt{T}\hat{\sigma}_y \Phi^{-1}(\tau) - \left(T\mu_y + \sqrt{T}\sigma_y \Phi^{-1}(\tau)\right), \\ a_n =& \left(T\psi_T(\tau)\sigma_y\right)/\sqrt{n}. \end{aligned}$$

By Lemma A5 we have

$$\widehat{Q}^{\text{boot}}_{Y_T}(\tau) - Q_{Y_T}(\tau) = W_n + R_n,$$

where $R_n = O\left(n^{-1/2} + T^{-1/2}\right)$. Lemma A1 implies that a similar expansion holds for the skewness-corrected MLE. In the following, we will only consider the bootstrap estimator. The proof for the skewness-corrected MLE is analogous.

Noting $\nu_T(\tau)/\psi_T(\tau) \to 1$, we have

$$a_n^{-1} R_n = O_p\left(T^{-1} + (T^3/n)^{-1/2}\right) + o_p(1).$$

Moreover, by Markov's inequality and the fact that the fourth moment of y_t is bounded,

$$\hat{\sigma}_y/\sigma_y - 1 = O_p\left(n^{-1/2}\right)$$

and hence

$$a_n^{-1}\left\{\frac{T}{\sqrt{n}}\psi_T(\tau)\sigma_y\left(\hat{\sigma}_y/\sigma_y - 1\right)\Phi^{-1}\left(1-\alpha/2\right)\right\} = O_p\left(n^{-1/2}\right).$$

Therefore, by Lemma A2,

$$a_n^{-1}\left(\widehat{Q}_{Y_T}^{\text{boot}}(\tau)-Q_{Y_T}(\tau)-\frac{T}{\sqrt{n}}\psi_T(\tau)\sigma_y\left(\hat{\sigma}_y/\sigma_y-1\right)\Phi^{-1}\left(1-\alpha/2\right)\right)\Rightarrow N(0,1).$$

Thus,

$$\begin{aligned}
&P\left(\widehat{Q}_{Y_T}^{\ell}(\tau;\alpha)>Q_{Y_T}(\tau)\right)\\
=&P\Bigg(\widehat{Q}_{Y_T}^{\text{boot}}(\tau)-Q_{Y_T}(\tau)-\frac{T}{\sqrt{n}}\psi_T(\tau)\sigma_y\left(\hat{\sigma}_y/\sigma_y-1\right)\Phi^{-1}\left(1-\alpha/2\right)\\
&\qquad>\frac{T}{\sqrt{n}}\psi_T(\tau)\sigma_y\Phi^{-1}\left(1-\alpha/2\right)\Bigg)\\
=&P\Bigg(a_n^{-1}\left(\widehat{Q}_{Y_T}^{\text{boot}}(\tau)-Q_{Y_T}(\tau)-\frac{T}{\sqrt{n}}\psi_T(\tau)\sigma_y\left(\hat{\sigma}_y/\sigma_y-1\right)\Phi^{-1}\left(1-\alpha/2\right)\right)\\
&\qquad>\Phi^{-1}\left(1-\alpha/2\right)\Bigg)\\
\to&P\left(N(0,1)>\Phi^{-1}\left(1-\alpha/2\right)\right)=\alpha/2.
\end{aligned}$$

In the same way, it can be shown than

$$P\left(\widehat{Q}_{Y_T}^{u}(\tau;\alpha)>Q_{Y_T}(\tau)\right)\to\alpha/2.$$

In summary,

$$\begin{aligned}
&P\left(Q_{Y_T}(\tau)\notin\left[\widehat{Q}_{Y_T}^{\ell}(\tau;\alpha),\widehat{Q}_{Y_T}^{u}(\tau;\alpha)\right]\right)\\
=&P\left(\widehat{Q}_{Y_T}^{\ell}(\tau;\alpha)>Q_{Y_T}(\tau)\right)+P\left(\widehat{Q}_{Y_T}^{u}(\tau;\alpha)<Q_{Y_T}(\tau)\right)\to\alpha/2+\alpha/2=\alpha.
\end{aligned}$$

□

A7.2 Supporting lemmas

Lemma A1 (Equivalence bootstrap and skewness-corrected MLE)**.** *Suppose that the distribution of y_t admits a density with respect to Lebesgue measure and that*

$$\mathbb{E}\left[|y_t|^8\right] < \infty.$$

Then,

$$\sup_{\epsilon \leq \tau \leq 1-\epsilon} \left| \widehat{Q}_{Y_T}^{boot}(\tau) - \widehat{Q}_{Y_T}^{ML\text{-}skew}(\tau) \right| = O_p\left((n/T)^{-1/2} + T^{-1/2}\right).$$

Proof. Plugging in the expansion from Lemma A4,

$$\begin{aligned} &\widehat{Q}_{Y_T}^{\text{ML-skew}}(\tau) - Q_{Y_T}(\tau) \\ =&T(\hat{\mu}_y - \mu_y) + \sqrt{T}(\hat{\sigma}_y - \sigma)\Phi^{-1}(\tau) + \frac{1}{6}\left(\hat{\sigma}_y\hat{\gamma}_y - \sigma_y\gamma_y\right)\left(\left(\Phi^{-1}(\tau)\right)^2 - 1\right). \end{aligned}$$

By standard arguments (see e.g. the proof of Lemma A2),

$$\hat{\sigma}_y\hat{\gamma}_y - \sigma_y\gamma_y = O_p\left(n^{-1/2}\right).$$

The conclusion now follows by Lemma A5. □

Lemma A2. *Suppose that there is $\epsilon > 0$ such that*

$$\mathbb{E}\left[\left(\frac{y_t - \mu_y}{\sigma_y}\right)^{4+\epsilon}\right] < \infty.$$

Let

$$\nu_T^2(\tau) = \mathbb{E}\left[\left(\frac{y_t - \mu_y}{\sigma_y} + \frac{1}{2\sqrt{T}}\left[\left(\frac{y_t - \mu_y}{\sigma_y}\right)^2 - 1\right]\Phi^{-1}(\tau)\right)^2\right].$$

Then $\nu_T(\tau)$ is bounded away from zero and infinity and

$$\begin{aligned}&\frac{\sqrt{n}}{T\sigma_y\nu_T(\tau)}\left(T\hat{\mu}_y+\sqrt{T}\hat{\sigma}_y\Phi^{-1}(\tau)-\left(T\mu_y+\sqrt{T}\sigma_y\Phi^{-1}(\tau)\right)\right)\\=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{y_t-\mu_y}{\sigma_y}+\frac{1}{2\sqrt{T}}\left[\left(\frac{y_t-\mu_y}{\sigma_y}\right)^2-1\right]\Phi^{-1}(\tau)\right\}/\nu_T(\tau)+O_p\left((nT)^{-1/2}\right).\end{aligned}$$

The right-hand side converges to a standard normal random variable as $n\to\infty$.

Proof of Lemma A2. For later reference, we prove an asymptotic expansion of $\frac{\hat{\sigma}_y}{\sigma_y}-1$. First, note that

$$\frac{\hat{\mu}_y-\mu_y}{\sigma_y}=O_p\left(n^{-1/2}\right)$$

and therefore

$$\begin{aligned}\hat{\sigma}_y^2/\sigma_y^2=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t-\hat{\mu}_y}{\sigma_y}\right)^2=&\frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t-\mu_y}{\sigma_y}\right)^2-\left(\frac{\hat{\mu}_y-\mu_y}{\sigma_y}\right)^2\\=&\frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t-\mu_y}{\sigma_y}\right)^2+O_p\left(n^{-1}\right).\end{aligned}\tag{A49}$$

Thus, by Markov's inequality and the fact that the fourth moment of $(y_t-\mu_y)/\sigma_y$ is bounded,

$$\frac{\hat{\sigma}_y^2}{\sigma_y^2}-1=O_p\left(n^{-1/2}\right).\tag{A50}$$

For $w>0$,

$$\sqrt{w}-1=\frac{1}{2}(w-1)+O((w-1)^2).$$

This implies the stochastic expansion

$$\hat{\sigma}_y/\sigma_y-1=\sqrt{\hat{\sigma}_y^2/\sigma_y^2}-1=\frac{1}{n}\sum_{t=1}^{n}\left\{\left(\frac{y_i-\mu_y}{\sigma_y}\right)^2-1\right\}/2+O_p\left(n^{-1}\right),$$

where the order of the remainder term follows from (A49) and (A50). Now,

$$\begin{aligned}
&T\hat{\mu}_y + \sqrt{T}\hat{\sigma}_y\Phi^{-1}(\tau) - \left(T\mu_y + \sqrt{T}\sigma_y\Phi^{-1}(\tau)\right) \\
=&\frac{T}{\sqrt{n}}\sigma_y\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{T}\frac{y_t-\mu_y}{\sigma_y} + \sqrt{\frac{n}{T}}\left(\frac{\hat{\sigma}_y}{\sigma_y}-1\right)\Phi^{-1}(\tau)\right) \\
=&\frac{T}{\sqrt{n}}\sigma_y\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{T}\left(\frac{y_t-\mu_y}{\sigma_y} + \frac{1}{2\sqrt{T}}\left[\left(\frac{y_t-\mu_y}{\sigma_y}\right)^2-1\right]\Phi^{-1}(\tau)\right) + O_p\left((nT)^{-1/2}\right)\right\},
\end{aligned}$$

where the last inequality follow from the stochastic expansion for $\hat{\sigma}_y/\sigma_y - 1$. We have

$$\nu_T^2(\tau) = 1 + \frac{1}{\sqrt{T}}\gamma_y\Phi^{-1}(\tau) + \frac{1}{4T}\left(\kappa_y - 1\right)\left(\Phi^{-1}(\tau)\right)^2$$

bounded away from zero and infinity uniformly in T and over τ bounded away from zero and one. This proves

$$\begin{aligned}
&\frac{\sqrt{n}}{T\sigma_y\nu_T(\tau)}\left(T\hat{\mu}_y + \sqrt{T}\hat{\sigma}_y\Phi(\tau)^{-1} - \left(T\mu_y + \sqrt{T}\sigma_y\Phi(\tau)^{-1}\right)\right) \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{y_t-\mu_y}{\sigma_y} + \frac{1}{2\sqrt{T}}\left[\left(\frac{y_t-\mu_y}{\sigma_y}\right)^2-1\right]\Phi^{-1}(\tau)\right\}/\nu_T(\tau) + O_p\left((nT)^{-1/2}\right).
\end{aligned}$$

The Lyapunov condition for the Lindeberg-Feller central limit theorem for triangular arrays is easily checked and

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{y_t-\mu_y}{\sigma_y} + \frac{1}{2\sqrt{T}}\left[\left(\frac{y_t-\mu_y}{\sigma_y}\right)^2-1\right]\Phi^{-1}(\tau)\right\}/\nu_T(\tau) \Rightarrow N(0,1).$$

□

Lemma A3 (Glivenko-Cantelli for Fama-French bootstrap)**.** *Suppose that* $T^2/n \to 0$. *Then, there are universal constants* C_1, C_2 *and* c *such that*

$$P\left(\sup_{a\in\mathbb{R}}\left|P\left(Y_T^* \le a \mid \mathcal{Y}_n\right) - P\left(Y_T \le a\right)\right| > C_1\frac{T}{\sqrt{n}}\sqrt{\log\frac{n}{T^2}}\right) \le C_2\left(\frac{T^2}{n}\right)^c.$$

Proof of Lemma A3. Below, we show that for functions $g : \mathbb{R} \to [0,1]$,

$$P\left(\left|\,\mathbb{E}\left[g(Y_T^*) \mid \mathcal{Y}_n\right] - \mathbb{E}\left[g(Y_T^*)\right]\,\right| > s/2\right) \leq 2e^{-\frac{s^2 n}{32T^2}}. \tag{A51}$$

In particular, this convergence result holds for $g(y) = \mathbf{1}\{y \leq a\}$ for fixed $a \in \mathbb{R}$. In order to extend this pointwise result to a uniform result, let $a_{T,1} < \ldots < a_{T,n_a}$ such that, for each $a \in [0,1]$, there exist $\hat{a}_{T,1}(a), \hat{a}_{T,2}(a) \in \{a_{T,1}, \ldots, a_{T,n_a}\}$ such that $\hat{a}_{1,T}(a) \leq a \leq \hat{a}_{2,T}(a)$ and

$$\left|\mathbb{E}[\mathbf{1}\{Y_T^* \leq a\}] - \mathbb{E}[\mathbf{1}\{Y_T^* \leq \hat{a}_{k,T}(a)\}]\right| \leq s/2$$

for $k = 1, 2$. Using standard arguments for proving Glivenko-Cantelli results, it can be shown that n_a can be taken to satisfy $n_a \leq \tilde{C}/s$ for a universal constant $\tilde{C}$. Note that, even though the mass points $a_{T,1}, \ldots, a_{T,n_a}$ depend on T, n_a does not.

On the set

$$\mathcal{A} = \bigcap_{k=1}^{n_a} \left\{\left|\,\mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a_k\} \mid \mathcal{Y}_n\right] - \mathbb{E}[\mathbf{1}\{Y_T^* \leq a_k\}]\right| \leq s/2\right\}$$

we have for all $a \in \mathbb{R}$

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\} \mid \mathcal{Y}_n\right] - \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\}\right] \leq &\left(\mathbb{E}\left[\mathbf{1}\{Y_{T,j}^* \leq \hat{a}_{2,T}(a)\}\right] - \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq \hat{a}_{2,T}(a)\}\right]\right) \\
&+ \left(\mathbb{E}\left[\mathbf{1}\{Y_T^* \leq \hat{a}_{2,T}(a)\}\right] - \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\}\right]\right) \\
\leq &s/2 + s/2 = s
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\} \mid \mathcal{Y}_n\right] - \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\}\right] \geq &\left(\mathbb{E}\left[\mathbf{1}\{Y_{T,j}^* \leq \hat{a}_{1,T}(a)\}\right] - \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq \hat{a}_{1,T}(a)\}\right]\right) \\
&+ \left(\mathbb{E}\left[\mathbf{1}\{Y_T^* \leq \hat{a}_{1,T}(a)\}\right] - \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\}\right]\right) \\
\leq &s/2 + s/2 = s.
\end{aligned}$$

Therefore, on $\mathcal{A}$,

$$\sup_{a\in\mathbb{R}} \left| \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\} \mid \mathcal{Y}_n\right] - \mathbb{E}\left[\mathbf{1}\left\{Y_T^* \leq a\right\}\right] \right| \leq s$$

and by (A51)

$$\begin{aligned}
&P\left(\sup_{a\in\mathbb{R}} \left| \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\} \mid \mathcal{Y}_n\right] - \mathbb{E}\left[\mathbf{1}\left\{Y_T^* \leq a\right\}\right] \right| \leq s\right) \\
\leq & P\mathcal{A}^{\mathsf{c}} \\
\leq & \sum_{k=1}^{n_a} P\left(\left| \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a_k\} \mid \mathcal{Y}_n\right] - \mathbb{E}[\mathbf{1}\{Y_T^* \leq a_k\}]\right| > s/2\right) \\
\leq & (\tilde{C}/s) e^{-\frac{s^2 n}{32T^2}} \\
= & \exp\left(\log \tilde{C} - \log s - s^2/32(n/T^2)\right).
\end{aligned}$$

Now, choosing C_1 large enough and $s = (C_1/2)\frac{T}{\sqrt{n}}\sqrt{\log \frac{n}{T^2}}$ yields

$$P\left(\sup_{a\in\mathbb{R}} \left| \mathbb{E}\left[\mathbf{1}\{Y_{T,j}^* \leq a\}\right] - \mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\}\right] \right| > (C_1/2)\frac{T}{\sqrt{n}}\sqrt{\log \frac{n}{T^2}}\right) \leq C_2 \left(\frac{T^2}{n}\right)^c,$$

for universal constants C_2 and $c > 0$. The conclusion follows by combining this inequality with a bound on the bias. In particular, it suffices to show

$$\sup_{a\in\mathbb{R}} \left|\mathbb{E}\left[\mathbf{1}\{Y_T^* \leq a\}\right] - P\left[\mathbf{1}\{Y_T \leq a\}\right]\right| \leq \frac{4T^2}{n} \leq (C_1/2)\frac{T}{\sqrt{n}}\sqrt{\log \frac{n}{T^2}}.$$

We show this by proving that, for all functions $g : \mathbb{R} \to [0, 1]$,

$$\left|\mathbb{E}\left[g(Y_T^*)\right] - \mathbb{E}\left[g(Y_T)\right]\right| \leq 4T^2/n. \tag{A52}$$

In preparation of the proof, we let $\mathcal{Y}_n = (y_1, \ldots, y_n)$ denote the vector of observed returns in the estimation sample. The bootstrapped long-run return can be written as

$$Y_T^* = \sum_{t=1}^{T}\sum_{i=1}^{n} \omega_{it}^* y_i,$$

where $\Omega^* = (\omega^*_{it})_{i=1,\ldots,n,t=1,\ldots,T}$ is a $n \times T$ random matrix of bootstrap weights that is independent of $\mathcal{Y}_n$ and has the properties that (1) each column $\omega^*_{\bullet t}$ has a multinomial distribution with parameters $(1, 1/n, \ldots, 1/n)$ and (2) columns $(\omega^*_{\bullet 1}, \ldots, \omega^*_{\bullet T})$ are independent. Let $\mathcal{D}_{n,T}$ denote the event

$$\mathcal{D}_{n,T} = \mathcal{D}_{n,T}(\Omega^*) = \{\text{none of the rows of } \Omega^* \text{ has a row sum greater than } 1\}.$$

The matrix Ω^* has exactly one one in each column and zeros in all other entries. All possible configurations are equally likely to occur. Computing the probability of $\mathcal{D}_{n,T}$ is therefore a counting exercise:

$$\begin{aligned} P\mathcal{D}_{n,T} =& \frac{\#\text{ valid configurations compatible with } \mathcal{D}_{n,T}}{\#\text{ valid configurations}} \\ =& \frac{n(n-1)\cdots(n-T+1)}{n^T}. \end{aligned}$$

Thus,

$$\begin{aligned} 0 \leq 1 - P\mathcal{D}_{n,T} =& 1 - \exp(\log P\mathcal{D}_{n,T}) \\ \leq& 1 - 1 - \log P\mathcal{D}_{n,T} \\ \leq& -\sum_{j=0}^{T-1} \log\left(1 - \frac{j}{n}\right) \\ \leq& -(T-1)\log\left(1 - \frac{T-1}{n}\right) \\ \leq& (T-1)\frac{2(T-1)/n}{2-(T-1)/n} \leq \frac{2T^2}{n} \end{aligned}$$

and therefore $|P\mathcal{D}^{\mathsf{c}}_{n,T}| \leq 2T^2/n$, where the second inequality follows from $e^x \geq 1 + x$ and the second-to-last inequality follows from the inequality

$$\log(1+x) \geq \frac{2x}{2+x} \qquad \text{for } x > -1. \tag{A53}$$

Let F_Ω denote the distribution of the random $n \times T$ matrix Ω^*. Let F_y^n denote the distribution of $\mathcal{Y}_n$. Note that $F_y^n = \bigtimes_{i=1}^n F_y$, where F_y is the distribution of the single-

period log return y. Let ι_T denote the T-vector of ones so that we can write $Y_T^* = \mathcal{Y}_n'\Omega^*\iota_T$. By Fubini's theorem,

$$\begin{aligned}\mathbb{E}\left[\mathcal{D}_{n,T}g(Y_T^*)\right] &= \mathbb{E}\left[\mathcal{D}_{n,T}(\Omega^*)g\left(\mathcal{Y}_n'\Omega^*\iota_T\right)\right] \\ &= \int \mathcal{D}_{n,T}(w)g(y'w\iota_T)\,dF_\Omega(w)\,dF_y^n(y) \\ &= \int_w \mathcal{D}_{n,T}(w)\left(\int_y g(y'w\iota_T)\,dF_y^n(y)\right)\,dF_\Omega(w).\end{aligned}$$

Now, consider the inner integral. If $\mathcal{D}_{n,T}(w) = 1$ we can take $w\iota_T$ to be a vector with ones in T distinct entries and zeros in all other entries. Therefore, the $w\iota_T$ sums T distinct elements of y. The elements in y are independent and each have marginal measure F_y, the same distribution as the population single-period return. Therefore, $y'w\iota_T$ sums up T independent population single-period returns and the inner integral computes $\mathbb{E}\left[g(Y_T)\right]$. Thus,

$$\begin{aligned}\mathbb{E}\left[g(Y_T^*)\right] - \mathbb{E}\left[g(Y_T)\right] &= \mathbb{E}\left[\mathcal{D}_{n,T}g(Y_T^*)\right] - \mathbb{E}\left[g(Y_T)\right] + \mathbb{E}\left[\mathcal{D}_{n,T}^{\mathsf{c}}g(Y_T^*)\right] \\ &= \left(P\mathcal{D}_{n,T} - 1\right)\mathbb{E}\left[g(Y_T)\right] + \mathbb{E}\left[\mathcal{D}_{n,T}^{\mathsf{c}}(w_{i,T})g(Y_T^*)\right].\end{aligned}$$

Thus,

$$\left|\,\mathbb{E}\left[g(Y_T^*)\right] - \mathbb{E}\left[g(Y_T)\right]\,\right| \leq 2P\mathcal{D}_{n,T}^{\mathsf{c}} \leq 4T^2/n,$$

where we used the fact that the absolute value of g is bounded by one. This proves (A52).

Finally, we prove (A51). We consider the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) = \mathbb{E}\left[g\left(\mathcal{Y}_n'\Omega^*\iota_T\right) \mid \mathcal{Y}_n = x\right] = \mathbb{E}\left[g\left(\sum_{i=1}^n \left(\Omega^*\iota_n\right)_i y_i\right) \mid \mathcal{Y}_n = x\right].$$

Consider the maximal change in $f(x)$ from changing the ith entry of x. Any such change is bounded by one since $0 \leq g \leq 1$. In the definition of f, we are integrating over $(\Omega^*\iota_n)_i$, where $(\Omega^*\iota_n)_i$ gives the number of times that observation i was picked by the bootstrap.

On the set $\{(\Omega^*\iota_n)_i = 0\}$, changing x_i will have no effect. We can therefore bound

$$f(x) - f(y) \leq \sum_{i=1}^{n} \big\{ \underbrace{1}_{\text{maximal change}} \times P((\Omega^*\iota_n)_i > 0) + \underbrace{0}_{\text{no change}} \times P((\Omega^*\iota_n)_i = 0) \big\} \mathbf{1}_{x_i \neq y_i}.$$

The event $\mathcal{E}_{i,T} = \{(\Omega^*\iota_n)_i = 0\}$ can also be written as

$$\mathcal{E}_{i,T} = \{\text{the } i\text{th row of the matrix } \Omega^* \text{ contains only zeros}\}.$$

Again, we can use the fact that all possible configurations of the matrix Ω^* are equally likely and turn the computation of the probability of $\mathcal{E}_{i,T}$ into a counting exercise:

$$P\mathcal{E}_{i,T} = \frac{\#\text{ configurations with } \mathcal{E}_{i,T}}{\#\text{ valid configurations}} = \frac{(n-1)^T}{n^T} = \left(\frac{n-1}{n}\right)^T.$$

Noting that $-1/n > -1$ for $n \geq 2$, we can use inequality (A53) and conclude

$$0 \geq \log\left(\frac{n-1}{n}\right)^T = \log P\mathcal{E}_{i,T} = T\log\left(1 - \frac{1}{n}\right) \geq -T\frac{2/n}{2 - 1/n}. \tag{A54}$$

The inequality $e^x \geq 1 + x$ implies

$$P\mathcal{E}_{i,T}^{\mathsf{c}} = 1 - P\mathcal{E}_{i,T} = 1 - e^{\log P\mathcal{E}_{i,T}} \leq 1 - 1 - \log P\mathcal{E}_{i,T} = -\log P\mathcal{E}_{i,T}.$$

Therefore, we can take

$$f(x) - f(y) \leq \sum_{i=1}^{n} c_i \mathbf{1}_{x_i \neq y_i}$$

with

$$c_i = -\log P\mathcal{E}_{i,T}.$$

The sandwich (A54) implies,

$$c_i^2 \leq 4(T/n)^2$$

This yields the bound

$$\sum_{i=1}^{n} c_i^2 \leq n4(T/n)^2 = 4T^2/n.$$

Now, Lemma A6 (Talagrand's inequality) implies,

$$P\left(\left|\,\mathbb{E}\left[g(Y_T^*) \mid \mathcal{Y}_n\right] - \mathbb{E}\left[g(Y_T^*)\right]\,\right| > s\right) = P\left(\left|f(\mathcal{Y}_n) - \mathbb{E}\left[f(\mathcal{Y}_n)\right]\,\right|\right) \leq 2e^{-\frac{s^2 n}{8T^2}},$$

proving inequality (A51). □

Lemma A4 (Expansion of long-run return)**.** *Assume* $\sigma_y > 0$*,*

$$\mathbb{E}\left[|y_t|^4\right] < \infty$$

and that y_t *admits a density with respect to Lebesgue measure. Let*

$$\gamma_y = \mathbb{E}\left[(y_t - \mu_y)^3\right]/\sigma_y^3$$

and

$$\tilde{Q}_{Y_T}(\tau) = T\mu_y + \sqrt{T}\sigma_y \Phi^{-1}(\tau) + \frac{1}{6}\sigma_y\gamma_y\left(\left(\Phi^{-1}(\tau)\right)^2 - 1\right).$$

Then, for $\epsilon > 0$*,*

$$\sup_{\epsilon \leq \tau \leq 1-\epsilon} \left|\tilde{Q}_{Y_T}(\tau) - Q_{Y_T}(\tau)\right| = O\left(T^{-1/2}\right).$$

Proof of Lemma A4. Let

$$S_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{y_t - \mu_y}{\sigma_y} \right).$$

Lemma A7 implies the Edgeworth expansion

$$P\left(S_T \leq x\right) = \Phi\left(x\right) - T^{-1/2} \frac{1}{6} \gamma \left(x^2 - 1\right) + O\left(T^{-1}\right),$$

where the big O terms is uniform in x. This Edgeworth expansion implies a Cornish-Fisher expansion for the quantile function of S_T. The argument is explained in Hall (1992, pages 68-70) and omitted here. Let $Q_{S_T}(\tau)$ denote the τ-quantile of S_T. The Cornish-Fisher expansion is given by

$$Q_{S_T}(\tau) = \Phi^{-1}\left(\tau\right) + T^{-1/2} \frac{1}{6} \gamma \left(\left(\Phi^{-1}(\tau)\right)^2 - 1 \right) + O\left(T^{-1}\right)$$

and holds uniformly on $\epsilon \leq \tau \leq 1 - \epsilon$. Now, noting that

$$Y_T = \mu_y + \sqrt{T}\sigma_y S_T$$

is strictly increasing in S_T, the monotone property of quantiles implies

$$Q_{Y_T} = \mu_y + \sqrt{T}\sigma_y Q_{S_T}$$

and hence the conclusion. □

Lemma A5 (Expansion of bootstrapped log return)**.** *Suppose that the distribution of y_t admits a density with respect to Lebesgue measure and that*

$$\mathbb{E}\left[|y_t|^8\right] < \infty.$$

Then,

$$\begin{aligned}&\sup_{\epsilon\leq\tau\leq1-\epsilon}\left|\widehat{Q}_{Y_T}^{boot}(\tau)-Q_{Y_T}(\tau)-T(\hat{\mu}_y-\mu_y)-\sqrt{T}\left(\hat{\sigma}_y-\sigma_y\right)\Phi^{-1}(\tau)\right|\\=&O_p\left(n^{-1/2}+T^{-1/2}\right).\end{aligned}$$

Proof of Lemma A5. Let P_n denote the empirical measure on $\mathcal{Y}_n$, i.e., the measure that assigns probability $1/n$ to each of the mass points $y_1,\ldots,y_n$. P_n has the distribution function

$$P_n(t)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{y_i\leq t\}.$$

Let $Q_{Y_T|\mathcal{Y}_n}$ denote the quantile function associated with the measure P_n^* conditional on the estimation sample $\mathcal{Y}_n$. For $\bar{\rho}>0$ and $c_1>0$, let $\mathcal{P}_{n,T}(c,\bar{\rho})$ denote the class of probability measures P such that

$$\mathbb{E}_P\left[|y|^4\right]\leq\bar{\rho}$$

and P satisfies the weak Cramér condition with parameter $(1,c_1,1)$, i.e.,

$$\sup_{|t|\geq1}|\phi_P(t)|\leq1-\frac{c_1}{t^2},$$

where ϕ_P is the characteristic function of P. There is c_1 such that

$$\Pr\left(P_n\notin\mathcal{P}_n(c_1,\bar{\rho})\right)\leq n^{-1}. \tag{A55}$$

To proof this claim, combine Proposition 3.2 (setting $b=1$ and $R=1$) and Proposition 3.3 in Song (2020) to conclude the existence of $c_1>0$ such that

$$\Pr\left(P_n\text{ does not satisfy weak Cramér with }(2,c_1,1)\right)\leq\exp\left(-\frac{c_1^2n}{2}\right).$$

Let $\bar{\rho}=(1/2)\max\{\mathbb{E}\left[|y_t|^4\right],\mathbb{E}\left[|y_t|^8\right]\}$. By Markov's inequality and independence of the

$(y_t)_{1\leq t\leq n}$,

$$\Pr\left(\left|\frac{1}{n}\sum_{t=1}^{n}\left(|y_t|^4-\mathbb{E}\left[|y_t|^4\right]\right)\right|>(1/2)\bar{\rho}\right)\leq\frac{\sum_{t=1}^{n}\mathbb{E}\left[|y_t|^8\right]}{(1/2)\bar{\rho}n^2}\leq n^{-1}$$

and therefore

$$\Pr\left(\mathbb{E}_{P_n^*}\left[|y|^4\right]>\bar{\rho}\right)=\Pr\left(\frac{1}{n}\sum_{t=1}^{n}|y_n|^4>\bar{\rho}\right)\leq n^{-1}.$$

This proves inequality (A55). We work conditionally on the estimation sample $\mathcal{Y}_n$ and suppose that $P_n^*\in\mathcal{P}_n^*(c_1,\bar{\rho})$. Let

$$\begin{aligned}\hat{\mu}_y=&\mathbb{E}_{P_n}[y]=\frac{1}{n}\sum_{t=1}^{n}y_t,\\ \hat{\sigma}_y^2=\mathrm{var}_{P_n}(y)=&\mathbb{E}_{P_n}\left[(y-\hat{\mu}_y)^2\right]=\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{\mu}_y)^2\end{aligned}$$

and let $(y_1^*,\ldots,y_T^*)$ denote a sample of size T from P_n^*. Define

$$S_T^*=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{y_t^*-\hat{\mu}_y}{\hat{\sigma}_y}$$

By Lemma A7, S_T^* admits an Edgeworth expansion if $P_n^*\in\mathcal{P}_n^*(c_1,\bar{\rho})$. In that case

$$\Pr\left(S_T^*\leq x\mid\mathcal{Y}_n\right)=\Phi(x)-T^{-1/2}\frac{1}{6}\frac{\mathbb{E}_{P_n^*}\left[(y-\hat{\mu}_y)^3\right]}{\hat{\sigma}_y^3}(x^2-1)+O\left(T^{-1}\right),$$

where the constant in the big O term depends on $\mathcal{P}_n^*$ but does not depend on x or the data $\mathcal{Y}_n$. Arguing as in in the proof of Lemma A4, this implies a conditional Cornish-Fisher expansion for the quantile function of S_T^* denoted by $Q_{S_T^*|\mathcal{Y}_n}$. In particular,

$$Q_{S_T^*|\mathcal{Y}_n}(\tau)=\Phi^{-1}(\tau)+T^{-1/2}\frac{1}{6}\frac{\mathbb{E}_{P_n^*}\left[(y-\hat{\mu}_y)^3\right]}{\hat{\sigma}_y^3}\left(\left(\Phi^{-1}(\tau)\right)^2-1\right)+O\left(T^{-1}\right),$$

where the remainder term is uniform in $\epsilon\leq\tau\leq1-\epsilon$ and $P_n\in\mathcal{P}_n^*$ and hence also uniform

in the data $\mathcal{Y}_n$. Note that

$$\hat{\gamma}_y = \frac{\mathbb{E}_{P_n^*}\left[(y-\hat{\mu}_y)^3\right]}{\hat{\sigma}_y^3}.$$

The previous display implies

$$Q_{Y_T^*|\mathcal{Y}_n}(\tau) = T\hat{\mu}_y + \sqrt{T}\hat{\sigma}_y\Phi^{-1}(\tau) + \frac{1}{6}\hat{\sigma}_y\hat{\gamma}_y\left(\left(\Phi^{-1}(\tau)\right)^2 - 1\right) + O\left(T^{-1/2}\right).$$

Therefore,

$$\begin{aligned} &Q_{Y_T^*|\mathcal{Y}_n}(\tau) - Q_{Y_T}(\tau) - T(\hat{\mu}_y - \mu_y) - \sqrt{T}(\hat{\sigma}_y - \sigma_y)\Phi^{-1}(\tau) \\ =&\frac{1}{6}\left(\hat{\sigma}_y\hat{\gamma}_y - \sigma_y\gamma_y\right)\left(\left(\Phi^{-1}(\tau)\right)^2 - 1\right) + O\left(T^{-1/2}\right). \end{aligned} \tag{A56}$$

By Markov's inequality, for $p = 1, 2, 3$ and every $\xi > 0$,

$$\Pr\left(\left|\mathbb{E}_{P_n}\left[(y-\hat{\mu}_y)^p\right] - \mathbb{E}\left[(y-\mu_y)^p\right]\right| > \xi\frac{\left(\mathbb{E}\left[|y-\mu_y|^{2p}\right]\right)^{1/2}}{\sqrt{n}}\right) \leq \xi^2.$$

Standard arguments and the inequality $|\sqrt{a}-1| \leq |a-1|$ imply

$$\begin{aligned} \hat{\sigma}_y/\sigma_y - 1 =& O_p\left(n^{-1/2}\right) \quad \text{and} \\ \hat{\gamma}_y - \gamma_y =& O_p\left(n^{-1/2}\right). \end{aligned}$$

Therefore,

$$\sup_{\epsilon\leq\tau\leq 1-\epsilon}\left|\frac{1}{6}\left(\hat{\sigma}_y\hat{\gamma}_y - \sigma_y\gamma_y\right)\left(\left(\Phi^{-1}(\tau)\right)^2 - 1\right)\right| = O_p\left(n^{-1/2}\right),$$

yielding the conclusion. □

Lemma A6 (Talagrand's inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables and suppose*

$$f(x) - f(y) \leq \sum_{i=1}^n c_i(x)\mathbf{1}_{x_i \neq y_i}$$

for all $x, y \in \mathbb{R}^n$. Then $f(X_1, \ldots, X_n)$ is $\|\sum_{i=1}^n c_i^2\|_\infty$-subgaussian. In particular,

$$P\left(|f(X_1, \ldots, X_n) - \mathbb{E}\left[f(X_1, \ldots, X_n)\right]| > s\right) \leq 2e^{-\frac{s^2}{2\|\sum_{i=1}^n c_i^2\|_\infty}}$$

for all $s > 0$.

Proof of Lemma A6. This statement of Talagrand's inequality and a proof can be found in van Handel (2016). □

Lemma A7 (Edgeworth expansion under weak Cramér condition)**.** *Let y denote a random variable with distribution $P \in \mathcal{P}_T$ and characteristic function ϕ_P and mean zero and variance one under all $P \in \mathcal{P}$. Suppose that there is $\bar{\rho} < \infty$ and $s \geq 3$ such that*

$$\sup_{T \geq 1} \sup_{P \in \mathcal{P}_T} \mathbb{E}_P\left[|y|^s\right] \leq \bar{\rho}.$$

In addition, suppose that y satisfies the uniform weak Cramér condition with parameter (b, c, R) such that $R \geq 1/(15\bar{\rho})$, $0 < b < 2/\max\{s-3, 1\}$ and $c > 0$, i.e., suppose that for these parameter values and $|t| \geq R$

$$\sup_{P \in \mathcal{P}_T} |\phi_P| \leq 1 - \frac{c}{|t|^b}.$$

Let $(y_1, \ldots, y_T)$ denote a vector of independent draws from y and

$$S_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t.$$

Then for all $T \geq 1$,

$$\sup_{P \in \mathcal{P}_T} \sup_{x \in \mathbb{R}} |P\left(S_T \leq x\right) - \Phi(x) - B_{P,s,T}(x)| = O\left(T^{-(s-2)/2}\right),$$

where Φ is the distribution function of a standard normal random variable and $B_{P,s,T}(x)$ is a polynomial in x of degree $3(s-3)$ with coefficients that depend on T and the cumulants

of y under P up to order $s-1$. In particular,

$$B_{P,T,4}(x) = -T^{-1/2}\frac{\mathbb{E}_P\left[y^3\right]}{6}\left(x^2-1\right).$$

Proof of Lemma A7. Apply Theorem 2.1 in Song (2020) to i.i.d. draws from the random variable y setting $f(y) = 1\{y \leq x\}$ for $x \in \mathbb{R}$, setting $n = T$. Note that for $Q_{n,P}$ in Theorem 2.1 in Song (2020),

$$P\left(S_T \leq x\right) = \int f\, dQ_{n,s,P}.$$

The right-hand side of the approximation bound in Theorem 2.1 in Song (2020) contains the terms $M_s(f)$ and $\bar{\omega}_f\left(n^{-(s-2)/2};\Phi\right)$. For our choice of f, $M_s(f) \leq 1$ and

$$\bar{\omega}_f\left(n^{-(s-2)/2};\Phi\right) = O\left(n^{-(s-2)/2}\right).$$

This bounds the approximation error by a constant times $n^{-(s-2)/2}$. The constant does not depend on x or P. The approximation term in Theorem 2.1 in Song (2020) can be written as

$$\int f\, d\tilde{Q}_{T,s,P} = \Phi(x) + B_{P,s,T}(x) + E_{P,s,T}(x),$$

where

$$E_{P,s,T}(x) = n^{-(s-2)/2} r_{P,s}(x)\phi(x),$$

ϕ is the standard normal density and $r_{P,s}$ is a polynomial of degree $3(s-2)$ with coefficients that depend on the cumulants of y of up to order s (see Hall, 1992, pages 44-45). Under the assumptions of the lemma, these cumulants are bounded by a constant that depends only on $\bar{\rho}$. We have

$$\lim_{|x|\to\infty} r_{P,s}(x)\phi(x) = 0.$$

Therefore, $r_{P,s}(x)\phi(x)$ is bounded uniformly over $x \in \mathbb{R}$ and $P \in \mathcal{P}_T$. This allows us to subsume $E_{P,s,T}$ in the $O(T^{-(s-2)/2})$ term. □

Lemma A8 (Song, 2020)**.** *Let $(y_1, \ldots, y_n)$ denote a sample of independent random variables from a distribution $P \in \mathcal{P}_n$. Suppose that there are constants $c_L < c_U$ and measurable maps f_P, $P \in \mathcal{P}_n$, such that*

$$\inf_{n \geq 1} \inf_{P \in \mathcal{P}_n} \inf_{y \in [c_L, c_U]} f_P > 0$$

and

$$P\left(y \in [c_L, c_U] \cap A\right) = \int_{[c_L, c_U] \cap A} f_P(y)\, dy$$

for every $P \in \mathcal{P}_n$ and every Borel set A. Let

$$P_n^* = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$$

denote the empirical measure induced by $(y_1, \ldots, y_n)$. Let $\mathcal{C}_n(b, c, R)$ denote the event

$$\mathcal{C}_n(b, c, R) = \{P_n^* \textit{ satisfies the weak Cramér condition with parameter } (b, c_1, R)\}.$$

To every $b > 0$ and $R > 0$ there exist $c_1, c_2 > 0$ such that

$$\sup_{P \in \mathcal{P}_n} P\mathcal{C}_n(b, c_1, R) \leq \exp\left(-c_2 n\right).$$

Proof of Lemma A8. This is straightforward corollary to Proposition 3.2 and Proposition 3.3 in Song (2020). □

References

Anarkulova, A., Cederburg, S., O'Doherty, M.S., 2022. Stocks for the long run? Evidence from a broad sample of developed markets. *Journal of Financial Economics* 143, 409-433.

Fama, E., and K. French, 2018. Long Horizon Returns. *The Review of Asset Pricing Studies* 8, 232–252.

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer Science & Business Media, New York.

Merton, R.C., 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323-361.

Song, K. (2020). A uniform-in-P Edgeworth expansion under weak Cramér conditions. *Statistics* 54, 926—950.

Van Handel, R. (2016). Probability in high dimension. *Lecture notes*, Princeton university.

Table A1: Panel-data simulation results.

The table shows simulation results based on panel-data estimates with a sample size of $n = 1,440$ and $K = 20$, and compounding horizons $T = 120$ (Panels A1 and B1) and $T = 360$ (Panels A2 and B2). Monthly period returns are generated according to a one-factor model with betas uniformly distributed between 0.7 and 1.3. Two different return models are considered, as described in the main text: i.i.d. log-normal (Panels A1 and A2) and i.i.d. log-normal-with-crashes (Panels B1 and B2). All specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. The common factor in returns account for 40% of the total variance ($\lambda = 0.4$). The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile). The next row gives the true population values of the distribution for those percentile values (i.e., the quantiles of the distribution). The true population distribution is defined as the distribution for an asset with $\beta = 1$. The following the three rows show the actual average coverage rates of 90% nominal confidence bands, centered on the (i) pooled ML estimate, (ii) pooled skewness-corrected ML estimate, and (iii) pooled FF bootstrap estimate. The subsequent three rows in each panel show the median error (median bias) for each of the three estimators, for a given percentile of the distribution. The final three rows in each panel show the corresponding median absolute errors (median absolute biases). The results are based on 10,000 simulated samples.

Panel A1: Log-normal, $T = 120$									
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.36	0.57	0.72	1.07	1.66	2.57	3.82	4.85	7.56
	Coverage rates								
ML	89%	90%	90%	89%	89%	89%	88%	88%	88%
ML-skew	89%	90%	90%	89%	89%	89%	88%	88%	88%
Bootstrap	90%	90%	90%	89%	89%	89%	88%	88%	88%
	Median error								
ML	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08
ML-skew	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08
Bootstrap	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.04	0.08
	Median absolute error								
ML	0.03	0.05	0.06	0.09	0.14	0.22	0.34	0.43	0.69
ML-skew	0.03	0.05	0.06	0.09	0.14	0.22	0.34	0.43	0.68
Bootstrap	0.03	0.05	0.06	0.09	0.14	0.22	0.34	0.43	0.69

Panel A2: Log-normal, $T = 360$									
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.33	0.71	1.07	2.12	4.55	9.75	19.36	29.20	63.09
	Coverage rates								
ML	90%	89%	89%	89%	89%	89%	89%	89%	88%
ML-skew	90%	89%	89%	89%	89%	89%	89%	89%	88%
Bootstrap	90%	90%	89%	89%	89%	89%	89%	89%	88%
	Median error								
ML	0.00	-0.01	0.00	0.00	0.02	0.08	0.23	0.43	1.22
ML-skew	0.00	-0.01	0.00	0.00	0.02	0.08	0.23	0.44	1.23
Bootstrap	0.00	-0.01	-0.01	0.00	0.02	0.08	0.23	0.45	1.27
	Median absolute error								
ML	0.08	0.18	0.27	0.54	1.16	2.50	4.97	7.54	16.42
ML-skew	0.08	0.18	0.27	0.54	1.16	2.50	4.97	7.55	16.42
Bootstrap	0.08	0.18	0.27	0.54	1.16	2.49	4.97	7.52	16.42

Panel B1: Log-normal-with-crashes, $T = 120$									
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.30	0.51	0.67	1.04	1.66	2.60	3.86	4.86	7.43
	Coverage rates								
ML	82%	88%	89%	89%	89%	88%	87%	85%	76%
ML-skew	87%	88%	89%	89%	89%	88%	87%	87%	85%
Bootstrap	87%	88%	89%	89%	89%	88%	87%	87%	85%
	Median error								
ML	0.03	0.02	0.01	-0.01	-0.02	0.00	0.10	0.22	0.69
ML-skew	0.00	0.00	0.00	0.00	0.01	0.03	0.05	0.06	0.03
Bootstrap	0.00	0.00	0.00	0.01	0.01	0.02	0.04	0.05	0.10
	Median absolute error								
ML	0.04	0.06	0.07	0.10	0.15	0.23	0.35	0.45	0.82
ML-skew	0.04	0.06	0.07	0.10	0.16	0.24	0.34	0.43	0.65
Bootstrap	0.04	0.06	0.07	0.10	0.15	0.23	0.34	0.43	0.66

Panel B2: Log-normal-with-crashes, $T = 360$									
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.25	0.59	0.93	1.97	4.41	9.73	19.55	29.50	63.27
	Coverage rates								
ML	88%	89%	89%	89%	89%	89%	88%	88%	86%
ML-skew	89%	89%	89%	89%	89%	89%	88%	88%	87%
Bootstrap	89%	89%	89%	89%	89%	89%	88%	88%	87%
	Median error								
ML	0.02	0.02	0.02	0.00	-0.03	0.07	0.69	1.70	7.03
ML-skew	0.00	0.00	0.00	0.02	0.06	0.17	0.43	0.69	1.26
Bootstrap	0.00	0.00	0.01	0.02	0.06	0.17	0.38	0.65	1.49
	Median absolute error								
ML	0.08	0.18	0.27	0.56	1.22	2.64	5.27	7.95	17.47
ML-skew	0.08	0.18	0.27	0.56	1.23	2.65	5.22	7.81	16.55
Bootstrap	0.08	0.18	0.27	0.56	1.23	2.65	5.22	7.84	16.66

Table A2: Block bootstrap simulation results for $T = 120$.

The table shows simulation results based on estimates with a sample size of $n = 1,440$ and a compounding horizon of $T = 120$. Monthly period returns are generated according to five different return models described in the main text and the Online Appendix: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals with log returns following an MA(60) process (Panel D); long-term reversals with log returns following an MA(120) process (Panel E). All five specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile). The next row gives the true population values of the distribution for those percentile values (i.e., the quantiles of the distribution). The following the three rows show the actual average coverage rates of 90% nominal confidence bands, centered on the (i) skewness-corrected ML estimate, (ii) block bootstrap estimate with a random block length with a mean of 60 months, and (iii) block bootstrap estimate with a fixed block length of 60 months. The subsequent three rows in each panel show the median error (median bias) for each of the three estimators, for a given percentile of the distribution. The final three rows in each panel show the corresponding median absolute errors (median absolute biases). The results are based on 10,000 simulated samples.

	Panel A: Log-normal								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.36	0.57	0.72	1.07	1.66	2.57	3.82	4.85	7.56
	Coverage rates								
ML-skew	90%	90%	90%	90%	90%	90%	90%	90%	90%
Block BS (120 rnd.)	67%	79%	83%	87%	89%	88%	84%	78%	67%
Block BS (60 fix)	73%	82%	85%	89%	90%	89%	85%	82%	73%
	Median error								
ML-skew	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	-0.01
Block BS (120 rnd.)	0.05	0.02	0.02	0.01	0.00	-0.03	-0.10	-0.20	-0.93
Block BS (60 fix)	0.03	0.02	0.01	0.01	0.00	-0.03	-0.07	-0.14	-0.50
	Median absolute error								
ML-skew	0.05	0.07	0.09	0.13	0.21	0.32	0.48	0.61	0.96
Block BS (120 rnd.)	0.08	0.10	0.11	0.14	0.21	0.35	0.58	0.81	1.58
Block BS (60 fix)	0.07	0.09	0.10	0.14	0.21	0.34	0.55	0.75	1.41

	Panel B: Log-normal-with-crashes								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.30	0.51	0.67	1.04	1.66	2.61	3.87	4.88	7.46
	Coverage rates								
ML-skew	80%	84%	85%	88%	90%	92%	93%	93%	94%
Block BS (120 rnd.)	58%	72%	79%	86%	89%	89%	87%	83%	73%
Block BS (60 fix)	63%	76%	81%	87%	90%	90%	88%	86%	79%
	Median error								
ML-skew	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	-0.05
Block BS (120 rnd.)	0.06	0.03	0.02	0.01	0.01	-0.02	-0.09	-0.19	-0.81
Block BS (60 fix)	0.03	0.02	0.02	0.01	0.00	-0.01	-0.05	-0.11	-0.45
	Median absolute error								
ML-skew	0.05	0.08	0.10	0.15	0.22	0.32	0.46	0.57	0.86
Block BS (120 rnd.)	0.09	0.11	0.12	0.15	0.23	0.35	0.57	0.78	1.46
Block BS (60 fix)	0.08	0.10	0.11	0.15	0.22	0.34	0.54	0.72	1.30

	Panel C: SV								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.30	0.52	0.69	1.07	1.70	2.64	3.84	4.78	7.12
	Coverage rates								
ML-skew	76%	85%	87%	88%	90%	91%	92%	92%	92%
Block BS (120 rnd.)	53%	69%	77%	85%	89%	90%	88%	84%	76%
Block BS (60 fix)	58%	72%	79%	86%	90%	91%	90%	88%	81%
	Median error								
ML-skew	0.05	0.03	0.02	0.00	-0.03	-0.04	0.01	0.09	0.41
Block BS (120 rnd.)	0.06	0.03	0.02	0.01	0.00	-0.03	-0.09	-0.19	-0.75
Block BS (60 fix)	0.04	0.02	0.02	0.01	0.00	-0.02	-0.07	-0.12	-0.40
	Median absolute error								
ML-skew	0.06	0.08	0.10	0.14	0.22	0.33	0.47	0.59	0.90
Block BS (120 rnd.)	0.09	0.11	0.12	0.16	0.22	0.34	0.53	0.72	1.29
Block BS (60 fix)	0.08	0.10	0.12	0.15	0.22	0.33	0.50	0.66	1.15

	Panel D: Long-term reversals (MA-60)								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.41	0.62	0.77	1.11	1.66	2.48	3.56	4.42	6.64
	Coverage rates								
ML-skew	85%	90%	91%	93%	93%	93%	91%	89%	85%
Block BS (120 rnd.)	75%	84%	88%	92%	92%	91%	88%	84%	75%
Block BS (60 fix)	78%	87%	89%	92%	93%	92%	89%	87%	79%
	Median error								
ML-skew	-0.05	-0.05	-0.05	-0.04	0.00	0.10	0.27	0.43	0.93
Block BS (120 rnd.)	0.04	0.02	0.01	0.01	0.00	0.00	-0.05	-0.11	-0.58
Block BS (60 fix)	0.01	0.00	0.00	0.00	0.00	0.01	0.00	-0.03	-0.21
	Median absolute error								
ML-skew	0.06	0.08	0.09	0.13	0.19	0.29	0.45	0.60	1.05
Block BS (120 rnd.)	0.08	0.09	0.10	0.14	0.19	0.30	0.47	0.65	1.19
Block BS (60 fix)	0.07	0.09	0.10	0.13	0.19	0.29	0.45	0.61	1.09

Table A2: Block bootstrap simulation results for $T = 120$ (continued).

	Panel E: Long-term reversals (MA-120)								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.40	0.61	0.76	1.10	1.66	2.50	3.62	4.52	6.85
	Coverage rates								
ML-skew	88%	91%	92%	93%	93%	93%	92%	91%	89%
Block BS (120 rnd.)	74%	83%	88%	91%	93%	91%	88%	83%	73%
Block BS (60 fix)	78%	86%	89%	92%	93%	92%	89%	86%	78%
	Median error								
ML-skew	-0.04	-0.04	-0.04	-0.03	0.00	0.07	0.19	0.31	0.69
Block BS (120 rnd.)	0.04	0.02	0.01	0.00	-0.01	-0.02	-0.05	-0.12	-0.62
Block BS (60 fix)	0.01	0.00	0.00	-0.01	-0.01	0.00	0.00	-0.02	-0.23
	Median absolute error								
ML-skew	0.05	0.08	0.09	0.13	0.19	0.29	0.44	0.57	0.95
Block BS (120 rnd.)	0.08	0.09	0.10	0.13	0.19	0.31	0.50	0.69	1.27
Block BS (60 fix)	0.07	0.09	0.10	0.13	0.19	0.30	0.48	0.65	1.18

Table A3: Block bootstrap simulation results for $T = 360$.

The table shows simulation results based on estimates with a sample size of $n = 1,440$ and a compounding horizon of $T = 360$. Monthly period returns are generated according to five different return models described in the main text and the Online Appendix: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals with log returns following an MA(60) process (Panel D); long-term reversals with log returns following an MA(120) process (Panel E). All five specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile). The next row gives the true population values of the distribution for those percentile values (i.e., the quantiles of the distribution). The following the three rows show the actual average coverage rates of 90% nominal confidence bands, centered on the (i) skewness-corrected ML estimate, (ii) block bootstrap estimate with a random block length with a mean of 60 months, and (iii) block bootstrap estimate with a fixed block length of 60 months. The subsequent three rows in each panel show the median error (median bias) for each of the three estimators, for a given percentile of the distribution. The final three rows in each panel show the corresponding median absolute errors (median absolute biases). The results are based on 10,000 simulated samples.

	Panel A: Log-normal								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.33	0.71	1.07	2.12	4.55	9.75	19.36	29.20	63.09
	Coverage rates								
ML-skew	90%	90%	90%	90%	90%	90%	90%	90%	90%
Block BS (120 rnd.)	79%	85%	87%	89%	90%	89%	87%	84%	79%
Block BS (60 fix)	85%	88%	89%	90%	90%	90%	89%	88%	85%
	Median error								
ML-skew	0.00	0.00	-0.01	-0.01	-0.03	-0.06	-0.10	-0.17	-0.25
Block BS (120 rnd.)	0.10	0.10	0.10	0.08	-0.01	-0.39	-1.82	-4.08	-14.89
Block BS (60 fix)	0.03	0.04	0.04	0.02	-0.01	-0.24	-0.87	-1.70	-5.69
	Median absolute error								
ML-skew	0.12	0.26	0.39	0.77	1.65	3.54	7.04	10.64	22.98
Block BS (120 rnd.)	0.17	0.31	0.43	0.80	1.67	3.61	7.58	11.85	28.04
Block BS (60 fix)	0.14	0.28	0.41	0.78	1.66	3.59	7.33	11.34	25.83

	Panel B: Log-normal-with-crashes								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.25	0.60	0.94	1.98	4.45	9.80	19.70	29.75	63.82
	Coverage rates								
ML-skew	85%	87%	87%	89%	90%	91%	92%	92%	93%
Block BS (120 rnd.)	74%	81%	84%	88%	90%	90%	88%	87%	83%
Block BS (60 fix)	79%	84%	86%	88%	90%	91%	90%	90%	88%
	Median error								
ML-skew	0.00	0.01	0.01	0.02	0.03	0.03	0.06	0.05	-0.34
Block BS (120 rnd.)	0.10	0.11	0.11	0.09	0.03	-0.29	-1.55	-3.61	-13.93
Block BS (60 fix)	0.04	0.05	0.05	0.05	0.02	-0.08	-0.57	-1.17	-4.07
	Median absolute error								
ML-skew	0.11	0.25	0.38	0.78	1.68	3.61	7.10	10.64	22.29
Block BS (120 rnd.)	0.15	0.30	0.42	0.80	1.68	3.70	7.74	12.08	27.65
Block BS (60 fix)	0.13	0.27	0.40	0.79	1.68	3.67	7.42	11.32	25.45

	Panel C: SV								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.26	0.64	1.00	2.10	4.66	10.10	19.84	29.49	61.13
	Coverage rates								
ML-skew	85%	87%	88%	89%	90%	90%	91%	91%	92%
Block BS (120 rnd.)	71%	79%	83%	87%	89%	90%	89%	88%	84%
Block BS (60 fix)	76%	82%	85%	88%	90%	91%	91%	91%	89%
	Median error								
ML-skew	0.05	0.06	0.05	0.02	-0.06	-0.21	-0.12	0.23	2.59
Block BS (120 rnd.)	0.11	0.12	0.11	0.09	0.02	-0.38	-1.58	-3.63	-12.84
Block BS (60 fix)	0.04	0.05	0.05	0.05	0.02	-0.17	-0.71	-1.37	-4.52
	Median absolute error								
ML-skew	0.12	0.26	0.40	0.81	1.75	3.72	7.20	10.65	22.00
Block BS (120 rnd.)	0.16	0.32	0.46	0.85	1.77	3.73	7.43	11.38	25.31
Block BS (60 fix)	0.14	0.29	0.43	0.84	1.76	3.69	7.15	10.72	22.88

	Panel D: Long-term reversals (MA-60)								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.43	0.85	1.23	2.29	4.55	9.04	16.78	24.29	48.62
	Coverage rates								
ML-skew	90%	92%	92%	93%	93%	93%	92%	92%	90%
Block BS (120 rnd.)	85%	89%	90%	93%	93%	93%	91%	89%	86%
Block BS (60 fix)	89%	91%	92%	93%	93%	93%	92%	91%	89%
	Median error								
ML-skew	-0.10	-0.14	-0.17	-0.16	0.00	0.72	2.62	4.93	14.57
Block BS (120 rnd.)	0.09	0.08	0.07	0.05	0.00	-0.19	-0.99	-2.28	-8.37
Block BS (60 fix)	0.00	-0.01	-0.02	-0.02	0.01	0.09	0.33	0.45	0.51
	Median absolute error								
ML-skew	0.15	0.29	0.41	0.75	1.53	3.08	5.96	8.99	19.68
Block BS (120 rnd.)	0.19	0.33	0.45	0.79	1.54	3.04	5.93	9.01	19.32
Block BS (60 fix)	0.16	0.30	0.43	0.77	1.53	3.05	5.82	8.57	18.05

Table A3: Block bootstrap simulation results for $T = 360$ (continued).

	Panel E: Long-term reversals (MA-120)								
Percentile	1%	5%	10%	25%	50%	75%	90%	95%	99%
Pop. value	0.42	0.84	1.22	2.27	4.55	9.10	16.99	24.70	49.81
	Coverage rates								
ML-skew	90%	92%	92%	93%	93%	93%	92%	92%	90%
Block BS (120 rnd.)	85%	89%	90%	92%	93%	92%	91%	89%	85%
Block BS (60 fix)	89%	91%	92%	93%	93%	93%	92%	91%	88%
	Median error								
ML-skew	-0.09	-0.14	-0.16	-0.17	-0.04	0.56	2.22	4.24	12.62
Block BS (120 rnd.)	0.07	0.06	0.05	0.02	-0.04	-0.27	-1.02	-2.25	-8.09
Block BS (60 fix)	-0.02	-0.04	-0.05	-0.06	-0.05	0.09	0.44	0.84	1.71
	Median absolute error								
ML-skew	0.15	0.29	0.41	0.76	1.53	3.11	6.02	9.00	19.41
Block BS (120 rnd.)	0.18	0.33	0.45	0.78	1.53	3.14	6.16	9.29	20.14
Block BS (60 fix)	0.16	0.30	0.42	0.76	1.53	3.12	6.01	9.00	19.47

Table A4: Countrywise estimates using the MLE, with $T = 10$ years

This table shows the point estimates and 90 percent confidence intervals (in parentheses) for the 10-year distributions of country-specific gross returns, for each country with a full history in the DMS data set. The estimates are based on data for the entire sample period from 1900 to 2020, using the MLE. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile).

	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
Australia	0.80	0.97	1.34	1.93	2.77	3.84	4.67
	(0.58-1.10)	(0.71-1.32)	(1.01-1.77)	(1.49-2.49)	(2.19-3.51)	(3.06-4.82)	(3.73-5.84)
Austria	0.24	0.34	0.59	1.09	2.02	3.51	4.89
	(0.14-0.41)	(0.20-0.56)	(0.37-0.93)	(0.71-1.68)	(1.32-3.07)	(2.29-5.38)	(3.15-7.59)
Belgium	0.40	0.52	0.80	1.30	2.12	3.27	4.25
	(0.27-0.59)	(0.35-0.76)	(0.56-1.15)	(0.93-1.83)	(1.52-2.95)	(2.34-4.57)	(3.03-5.96)
Canada	0.75	0.91	1.24	1.74	2.46	3.35	4.03
	(0.57-1.00)	(0.69-1.19)	(0.96-1.59)	(1.37-2.22)	(1.95-3.10)	(2.66-4.21)	(3.20-5.07)
Denmark	0.67	0.82	1.18	1.75	2.60	3.71	4.59
	(0.49-0.91)	(0.61-1.11)	(0.89-1.56)	(1.33-2.31)	(1.95-3.45)	(2.74-5.02)	(3.34-6.30)
Finland	0.42	0.58	0.97	1.72	3.04	5.10	6.94
	(0.26-0.69)	(0.36-0.92)	(0.63-1.49)	(1.15-2.56)	(2.06-4.50)	(3.42-7.59)	(4.61-10.44)
France	0.45	0.57	0.87	1.39	2.22	3.37	4.33
	(0.31-0.64)	(0.40-0.82)	(0.62-1.22)	(1.00-1.93)	(1.61-3.06)	(2.44-4.66)	(3.12-6.01)
Germany	0.25	0.36	0.68	1.39	2.82	5.32	7.79
	(0.09-0.68)	(0.15-0.87)	(0.35-1.35)	(0.85-2.28)	(1.89-4.20)	(3.42-8.29)	(4.64-13.09)
Ireland	0.47	0.61	0.94	1.53	2.49	3.86	5.02
	(0.29-0.76)	(0.39-0.95)	(0.64-1.39)	(1.09-2.16)	(1.82-3.41)	(2.84-5.26)	(3.66-6.89)
Italy	0.29	0.39	0.68	1.23	2.24	3.85	5.32
	(0.16-0.51)	(0.23-0.67)	(0.42-1.09)	(0.81-1.88)	(1.52-3.31)	(2.61-5.67)	(3.58-7.91)
Japan	0.29	0.42	0.77	1.52	2.96	5.42	7.79
	(0.13-0.68)	(0.20-0.89)	(0.42-1.41)	(0.95-2.43)	(2.01-4.37)	(3.69-7.98)	(5.12-11.83)
Netherlands	0.59	0.74	1.08	1.64	2.50	3.64	4.56
	(0.42-0.84)	(0.53-1.04)	(0.80-1.47)	(1.23-2.20)	(1.87-3.34)	(2.70-4.90)	(3.35-6.20)
New Zealand	0.74	0.91	1.28	1.87	2.74	3.86	4.74
	(0.51-1.07)	(0.65-1.28)	(0.95-1.72)	(1.43-2.45)	(2.12-3.54)	(2.95-5.05)	(3.56-6.30)
Norway	0.47	0.61	0.94	1.53	2.49	3.86	5.03
	(0.31-0.70)	(0.42-0.88)	(0.66-1.34)	(1.09-2.15)	(1.75-3.55)	(2.65-5.64)	(3.37-7.50)
Portugal	0.29	0.42	0.75	1.44	2.76	4.96	7.05
	(0.16-0.54)	(0.24-0.74)	(0.45-1.24)	(0.91-2.27)	(1.78-4.29)	(3.14-7.85)	(4.35-11.40)
SouthAfrica	0.71	0.89	1.30	1.98	3.01	4.39	5.51
	(0.50-1.00)	(0.64-1.24)	(0.96-1.77)	(1.47-2.66)	(2.24-4.04)	(3.23-5.97)	(4.00-7.58)
Spain	0.49	0.62	0.91	1.41	2.17	3.20	4.03
	(0.35-0.69)	(0.45-0.86)	(0.67-1.24)	(1.04-1.90)	(1.60-2.93)	(2.34-4.37)	(2.93-5.56)
Sweden	0.63	0.80	1.17	1.80	2.77	4.07	5.13
	(0.44-0.91)	(0.56-1.13)	(0.85-1.62)	(1.33-2.43)	(2.08-3.68)	(3.07-5.39)	(3.87-6.80)
Switzerland	0.60	0.75	1.06	1.57	2.32	3.31	4.08
	(0.44-0.83)	(0.55-1.01)	(0.80-1.41)	(1.19-2.07)	(1.77-3.04)	(2.52-4.34)	(3.10-5.38)
UK	0.64	0.79	1.13	1.69	2.52	3.61	4.48
	(0.44-0.93)	(0.56-1.12)	(0.83-1.55)	(1.28-2.24)	(1.93-3.29)	(2.75-4.74)	(3.38-5.94)
US	0.69	0.86	1.25	1.90	2.87	4.16	5.20
	(0.48-0.99)	(0.61-1.22)	(0.91-1.72)	(1.42-2.53)	(2.19-3.76)	(3.21-5.40)	(4.03-6.72)

Table A5: Countrywise estimates using the MLE, with $T = 30$ years

This table shows the point estimates and 90 percent confidence intervals (in parentheses) for the 30-year distributions of country-specific gross returns, for each country with a full history in the DMS data set. The estimates are based on data for the entire sample period from 1900 to 2020, using the MLE. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile).

	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
Australia	1.55	2.17	3.82	7.16	13.42	23.61	33.12
	(0.65-3.71)	(0.93-5.07)	(1.71-8.54)	(3.34-15.35)	(6.48-27.78)	(11.69-47.71)	(16.57-66.17)
Austria	0.10	0.17	0.45	1.29	3.76	9.81	17.41
	(0.02-0.40)	(0.04-0.69)	(0.12-1.70)	(0.35-4.72)	(1.06-13.34)	(2.78-34.59)	(4.92-61.66)
Belgium	0.28	0.45	0.95	2.21	5.12	10.90	17.14
	(0.09-0.85)	(0.15-1.31)	(0.34-2.71)	(0.80-6.13)	(1.88-13.96)	(4.03-29.53)	(6.33-46.41)
Canada	1.24	1.71	2.91	5.28	9.58	16.37	22.55
	(0.56-2.72)	(0.79-3.68)	(1.38-6.14)	(2.57-10.88)	(4.73-19.40)	(8.17-32.78)	(11.31-44.98)
Denmark	1.01	1.46	2.70	5.35	10.61	19.65	28.42
	(0.42-2.41)	(0.62-3.42)	(1.17-6.23)	(2.33-12.29)	(4.58-24.56)	(8.33-46.36)	(11.86-68.15)
Finland	0.45	0.77	1.87	5.06	13.64	33.31	56.85
	(0.12-1.71)	(0.21-2.81)	(0.54-6.52)	(1.52-16.87)	(4.19-44.39)	(10.30-107.72)	(17.52-184.42)
France	0.38	0.58	1.20	2.69	6.03	12.46	19.25
	(0.13-1.06)	(0.21-1.61)	(0.44-3.25)	(1.01-7.17)	(2.29-15.89)	(4.75-32.73)	(7.33-50.57)
Germany	0.13	0.26	0.79	2.67	9.11	27.45	53.12
	(0.01-1.36)	(0.03-2.15)	(0.13-4.71)	(0.60-11.85)	(2.55-32.53)	(8.32-90.58)	(15.84-178.08)
Ireland	0.46	0.73	1.55	3.60	8.36	17.87	28.14
	(0.13-1.59)	(0.22-2.37)	(0.51-4.66)	(1.29-10.02)	(3.18-21.97)	(7.02-45.44)	(11.16-70.90)
Italy	0.15	0.26	0.66	1.87	5.28	13.45	23.55
	(0.03-0.67)	(0.06-1.10)	(0.17-2.54)	(0.53-6.60)	(1.59-17.52)	(4.19-43.18)	(7.40-74.90)
Japan	0.20	0.38	1.09	3.48	11.12	31.67	59.23
	(0.03-1.50)	(0.06-2.43)	(0.21-5.54)	(0.85-14.27)	(3.20-38.63)	(9.98-100.53)	(19.04-184.24)
Netherlands	0.76	1.13	2.16	4.45	9.18	17.61	26.00
	(0.29-1.98)	(0.44-2.86)	(0.88-5.32)	(1.85-10.72)	(3.85-21.88)	(7.37-42.07)	(10.81-62.52)
New Zealand	1.32	1.88	3.40	6.57	12.71	22.99	32.78
	(0.51-3.39)	(0.76-4.64)	(1.46-7.93)	(2.95-14.63)	(5.85-27.58)	(10.61-49.79)	(15.01-71.58)
Norway	0.46	0.72	1.54	3.59	8.35	17.84	28.12
	(0.15-1.36)	(0.25-2.09)	(0.55-4.34)	(1.29-10.00)	(2.96-23.52)	(6.16-51.70)	(9.47-83.45)
Portugal	0.19	0.35	0.97	2.99	9.23	25.47	46.75
	(0.04-0.94)	(0.08-1.62)	(0.23-4.09)	(0.76-11.75)	(2.44-34.91)	(6.76-95.88)	(12.28-177.98)
South Africa	1.32	1.95	3.75	7.75	16.02	30.82	45.59
	(0.51-3.38)	(0.77-4.91)	(1.52-9.20)	(3.21-18.72)	(6.65-38.61)	(12.68-74.92)	(18.56-111.98)
Spain	0.45	0.67	1.31	2.78	5.87	11.52	17.24
	(0.17-1.16)	(0.26-1.71)	(0.52-3.29)	(1.12-6.89)	(2.37-14.55)	(4.62-28.72)	(6.87-43.29)
Sweden	0.95	1.42	2.77	5.83	12.27	23.98	35.81
	(0.35-2.59)	(0.53-3.77)	(1.08-7.09)	(2.36-14.39)	(5.11-29.44)	(10.17-56.50)	(15.31-83.75)
Switzerland	0.74	1.07	1.96	3.87	7.63	14.06	20.26
	(0.31-1.78)	(0.45-2.53)	(0.85-4.56)	(1.70-8.82)	(3.38-17.21)	(6.25-31.59)	(9.01-45.56)
UK	0.89	1.30	2.41	4.83	9.64	17.98	26.11
	(0.33-2.38)	(0.50-3.33)	(0.99-5.87)	(2.08-11.19)	(4.29-21.69)	(8.08-40.01)	(11.72-58.17)
US	1.18	1.74	3.32	6.81	13.95	26.60	39.14
	(0.44-3.17)	(0.67-4.53)	(1.33-8.28)	(2.85-16.27)	(6.05-32.13)	(11.86-59.65)	(17.69-86.62)

Table A6: Countrywise estimates using the skewness-corrected MLE, with $T = 10$ years

This table shows the point estimates and 90 percent confidence intervals (in parentheses) for the 10-year distributions of country-specific gross returns, for each country with a full history in the DMS data set. The estimates are based on data for the entire sample period from 1900 to 2020, using the skewness-corrected MLE. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile).

	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
Australia	0.77	0.95	1.36	1.97	2.80	3.78	4.49
	(0.55-1.06)	(0.70-1.30)	(1.03-1.80)	(1.53-2.54)	(2.22-3.55)	(3.01-4.74)	(3.58-5.62)
Austria	0.23	0.33	0.60	1.11	2.04	3.46	4.71
	(0.14-0.40)	(0.20-0.55)	(0.38-0.95)	(0.72-1.72)	(1.34-3.11)	(2.26-5.30)	(3.03-7.30)
Belgium	0.39	0.51	0.81	1.32	2.13	3.25	4.16
	(0.26-0.58)	(0.35-0.75)	(0.57-1.15)	(0.94-1.85)	(1.53-2.97)	(2.33-4.53)	(2.96-5.83)
Canada	0.74	0.90	1.24	1.76	2.47	3.32	3.94
	(0.56-0.98)	(0.69-1.18)	(0.96-1.60)	(1.39-2.24)	(1.96-3.12)	(2.64-4.17)	(3.13-4.96)
Denmark	0.67	0.82	1.18	1.75	2.60	3.71	4.59
	(0.49-0.91)	(0.61-1.12)	(0.89-1.56)	(1.32-2.31)	(1.95-3.45)	(2.74-5.02)	(3.34-6.31)
Finland	0.41	0.57	0.98	1.75	3.08	5.03	6.70
	(0.25-0.67)	(0.36-0.91)	(0.64-1.50)	(1.17-2.62)	(2.08-4.55)	(3.38-7.49)	(4.46-10.09)
France	0.44	0.57	0.88	1.40	2.23	3.35	4.27
	(0.31-0.63)	(0.40-0.81)	(0.63-1.23)	(1.01-1.94)	(1.61-3.07)	(2.43-4.63)	(3.08-5.92)
Germany	0.19	0.33	0.75	1.63	3.07	4.80	5.92
	(0.07-0.51)	(0.14-0.79)	(0.38-1.48)	(0.99-2.68)	(2.06-4.59)	(3.09-7.47)	(3.53-9.94)
Ireland	0.44	0.59	0.96	1.60	2.55	3.76	4.68
	(0.27-0.71)	(0.38-0.92)	(0.65-1.42)	(1.14-2.25)	(1.87-3.49)	(2.76-5.12)	(3.41-6.42)
Italy	0.26	0.38	0.69	1.29	2.30	3.74	4.93
	(0.15-0.47)	(0.22-0.65)	(0.43-1.11)	(0.85-1.96)	(1.56-3.40)	(2.54-5.51)	(3.31-7.32)
Japan	0.24	0.39	0.83	1.70	3.16	5.03	6.37
	(0.11-0.55)	(0.19-0.82)	(0.45-1.51)	(1.06-2.73)	(2.14-4.66)	(3.42-7.40)	(4.19-9.68)
Netherlands	0.58	0.74	1.09	1.66	2.51	3.62	4.48
	(0.41-0.83)	(0.53-1.03)	(0.80-1.48)	(1.24-2.23)	(1.88-3.36)	(2.68-4.87)	(3.29-6.10)
New Zealand	0.71	0.90	1.30	1.91	2.77	3.81	4.56
	(0.49-1.03)	(0.64-1.26)	(0.96-1.74)	(1.47-2.50)	(2.14-3.59)	(2.91-4.98)	(3.43-6.07)
Norway	0.47	0.61	0.94	1.53	2.49	3.87	5.03
	(0.31-0.70)	(0.42-0.89)	(0.66-1.34)	(1.09-2.15)	(1.75-3.55)	(2.65-5.65)	(3.37-7.51)
Portugal	0.28	0.41	0.77	1.49	2.81	4.86	6.65
	(0.15-0.51)	(0.23-0.72)	(0.46-1.27)	(0.94-2.35)	(1.81-4.37)	(3.07-7.68)	(4.11-10.77)
South Africa	0.70	0.89	1.30	1.99	3.02	4.37	5.45
	(0.50-0.99)	(0.64-1.23)	(0.96-1.77)	(1.48-2.67)	(2.25-4.06)	(3.22-5.95)	(3.96-7.50)
Spain	0.49	0.62	0.91	1.41	2.17	3.19	4.01
	(0.35-0.68)	(0.45-0.85)	(0.67-1.25)	(1.04-1.91)	(1.60-2.94)	(2.33-4.36)	(2.91-5.52)
Sweden	0.61	0.79	1.18	1.84	2.80	4.02	4.96
	(0.42-0.88)	(0.55-1.11)	(0.86-1.64)	(1.36-2.48)	(2.10-3.72)	(3.04-5.33)	(3.75-6.58)
Switzerland	0.59	0.74	1.07	1.58	2.33	3.29	4.02
	(0.43-0.81)	(0.55-1.00)	(0.80-1.42)	(1.20-2.08)	(1.78-3.06)	(2.51-4.31)	(3.05-5.30)
UK	0.61	0.78	1.15	1.73	2.55	3.56	4.30
	(0.42-0.89)	(0.55-1.10)	(0.84-1.57)	(1.31-2.29)	(1.95-3.33)	(2.71-4.67)	(3.25-5.70)
US	0.66	0.85	1.27	1.94	2.91	4.10	5.00
	(0.46-0.95)	(0.60-1.20)	(0.93-1.74)	(1.45-2.59)	(2.22-3.81)	(3.16-5.31)	(3.87-6.45)

Table A7: Countrywise estimates using the skewness-corrected MLE, with $T = 30$ years

This table shows the point estimates and 90 percent confidence intervals (in parentheses) for the 30-year distributions of country-specific gross returns, for each country with a full history in the DMS data set. The estimates are based on data for the entire sample period from 1900 to 2020, using the skewness-corrected MLE. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile).

	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
Australia	1.49	2.14	3.87	7.33	13.59	23.26	31.83
	(0.62-3.57)	(0.92-4.99)	(1.73-8.65)	(3.42-15.71)	(6.56-28.13)	(11.51-47.00)	(15.93-63.60)
Austria	0.09	0.17	0.45	1.32	3.80	9.67	16.76
	(0.02-0.39)	(0.04-0.68)	(0.12-1.72)	(0.36-4.83)	(1.07-13.51)	(2.74-34.10)	(4.73-59.36)
Belgium	0.28	0.44	0.96	2.24	5.15	10.81	16.77
	(0.09-0.84)	(0.15-1.30)	(0.34-2.73)	(0.81-6.21)	(1.89-14.06)	(3.99-29.28)	(6.19-45.39)
Canada	1.21	1.69	2.93	5.35	9.64	16.24	22.10
	(0.55-2.66)	(0.78-3.65)	(1.39-6.18)	(2.60-11.02)	(4.76-19.52)	(8.11-32.53)	(11.08-44.07)
Denmark	1.01	1.46	2.70	5.35	10.61	19.66	28.44
	(0.42-2.41)	(0.62-3.43)	(1.17-6.23)	(2.33-12.28)	(4.58-24.56)	(8.33-46.37)	(11.86-68.19)
Finland	0.43	0.76	1.90	5.16	13.79	32.88	54.92
	(0.11-1.65)	(0.21-2.77)	(0.54-6.59)	(1.55-17.21)	(4.24-44.88)	(10.17-106.33)	(16.93-178.16)
France	0.37	0.58	1.21	2.71	6.06	12.40	18.98
	(0.13-1.04)	(0.21-1.60)	(0.44-3.27)	(1.02-7.23)	(2.30-15.97)	(4.72-32.55)	(7.22-49.84)
Germany	0.10	0.24	0.86	3.14	9.95	24.75	40.36
	(0.01-1.03)	(0.03-1.93)	(0.14-5.14)	(0.71-13.92)	(2.79-35.51)	(7.50-81.68)	(12.04-135.32)
Ireland	0.43	0.71	1.58	3.75	8.56	17.40	26.21
	(0.12-1.48)	(0.22-2.31)	(0.53-4.77)	(1.35-10.45)	(3.26-22.47)	(6.84-44.25)	(10.40-66.06)
Italy	0.14	0.25	0.68	1.95	5.41	13.07	21.81
	(0.03-0.62)	(0.06-1.06)	(0.18-2.61)	(0.55-6.90)	(1.63-17.96)	(4.07-41.95)	(6.86-69.37)
Japan	0.17	0.35	1.16	3.91	11.86	29.37	48.49
	(0.02-1.23)	(0.06-2.25)	(0.23-5.90)	(0.95-16.05)	(3.41-41.19)	(9.25-93.23)	(15.59-150.82)
Netherlands	0.75	1.12	2.17	4.50	9.23	17.49	25.56
	(0.29-1.95)	(0.44-2.84)	(0.88-5.35)	(1.87-10.83)	(3.87-22.00)	(7.32-41.80)	(10.63-61.47)
New Zealand	1.27	1.85	3.44	6.72	12.86	22.67	31.58
	(0.49-3.27)	(0.75-4.58)	(1.48-8.03)	(3.02-14.95)	(5.92-27.91)	(10.46-49.10)	(14.46-68.96)
Norway	0.46	0.72	1.54	3.59	8.35	17.85	28.13
	(0.15-1.36)	(0.25-2.09)	(0.55-4.34)	(1.29-10.00)	(2.96-23.52)	(6.16-51.71)	(9.48-83.48)
Portugal	0.18	0.34	0.99	3.09	9.40	24.92	44.14
	(0.04-0.89)	(0.07-1.58)	(0.23-4.16)	(0.79-12.16)	(2.49-35.56)	(6.62-93.83)	(11.59-168.03)
South Africa	1.30	1.94	3.76	7.79	16.07	30.70	45.13
	(0.51-3.35)	(0.77-4.89)	(1.53-9.23)	(3.22-18.83)	(6.67-38.73)	(12.63-74.64)	(18.37-110.87)
Spain	0.44	0.67	1.31	2.79	5.88	11.49	17.13
	(0.17-1.15)	(0.26-1.70)	(0.52-3.30)	(1.12-6.92)	(2.37-14.58)	(4.61-28.65)	(6.82-43.01)
Sweden	0.92	1.40	2.80	5.94	12.40	23.67	34.62
	(0.34-2.50)	(0.53-3.72)	(1.09-7.16)	(2.41-14.68)	(5.17-29.76)	(10.05-55.79)	(14.80-80.98)
Switzerland	0.73	1.06	1.97	3.90	7.67	13.98	19.96
	(0.30-1.75)	(0.45-2.51)	(0.85-4.58)	(1.71-8.90)	(3.40-17.29)	(6.22-31.41)	(8.88-44.89)
UK	0.86	1.28	2.45	4.94	9.77	17.71	25.08
	(0.32-2.29)	(0.50-3.28)	(1.01-5.95)	(2.13-11.45)	(4.34-21.97)	(7.96-39.41)	(11.26-55.87)
US	1.14	1.72	3.37	6.97	14.13	26.20	37.58
	(0.43-3.04)	(0.66-4.46)	(1.35-8.39)	(2.92-16.66)	(6.13-32.55)	(11.68-58.74)	(16.98-83.17)

Table A8: Countrywise estimates using the FF bootstrap estimator, with $T = 10$ years

This table shows the point estimates and 90 percent confidence intervals (in parentheses) for the 10-year distributions of country-specific gross returns, for each country with a full history in the DMS data set. The estimates are based on data for the entire sample period from 1900 to 2020, using the FF bootstrap estimator. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile).

	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
Australia	0.76	0.95	1.36	1.98	2.81	3.79	4.49
	(0.55-1.06)	(0.70-1.30)	(1.03-1.80)	(1.54-2.55)	(2.22-3.56)	(3.02-4.75)	(3.59-5.63)
Austria	0.23	0.34	0.61	1.12	2.03	3.45	4.76
	(0.14-0.40)	(0.20-0.56)	(0.38-0.96)	(0.73-1.73)	(1.34-3.09)	(2.25-5.29)	(3.07-7.39)
Belgium	0.39	0.52	0.82	1.34	2.16	3.28	4.21
	(0.27-0.59)	(0.36-0.76)	(0.57-1.17)	(0.95-1.88)	(1.55-3.01)	(2.35-4.58)	(3.00-5.90)
Canada	0.73	0.9	1.24	1.76	2.48	3.33	3.95
	(0.55-0.97)	(0.68-1.18)	(0.96-1.60)	(1.39-2.24)	(1.96-3.12)	(2.65-4.19)	(3.14-4.97)
Denmark	0.66	0.82	1.17	1.72	2.52	3.61	4.51
	(0.48-0.90)	(0.61-1.11)	(0.88-1.55)	(1.30-2.27)	(1.90-3.36)	(2.67-4.89)	(3.29-6.20)
Finland	0.41	0.57	0.98	1.74	3.03	4.95	6.65
	(0.25-0.66)	(0.36-0.91)	(0.64-1.51)	(1.17-2.60)	(2.05-4.48)	(3.32-7.36)	(4.42-10.01)
France	0.44	0.57	0.88	1.41	2.25	3.38	4.32
	(0.31-0.63)	(0.40-0.81)	(0.63-1.23)	(1.02-1.96)	(1.63-3.10)	(2.45-4.67)	(3.12-5.99)
Germany	0.17	0.4	0.84	1.53	2.69	4.51	6.19
	(0.06-0.47)	(0.16-0.96)	(0.43-1.67)	(0.93-2.52)	(1.81-4.02)	(2.90-7.02)	(3.69-10.40)
Ireland	0.43	0.6	0.98	1.59	2.5	3.72	4.69
	(0.27-0.70)	(0.38-0.93)	(0.66-1.44)	(1.13-2.24)	(1.83-3.42)	(2.73-5.06)	(3.42-6.44)
Italy	0.27	0.39	0.71	1.29	2.27	3.7	4.95
	(0.15-0.47)	(0.23-0.67)	(0.44-1.14)	(0.85-1.97)	(1.53-3.35)	(2.51-5.45)	(3.33-7.36)
Japan	0.24	0.41	0.88	1.67	2.95	4.8	6.4
	(0.10-0.55)	(0.20-0.87)	(0.48-1.60)	(1.05-2.68)	(2.00-4.35)	(3.26-7.06)	(4.21-9.73)
Netherlands	0.58	0.74	1.09	1.66	2.5	3.6	4.5
	(0.41-0.82)	(0.53-1.03)	(0.80-1.49)	(1.24-2.23)	(1.87-3.34)	(2.67-4.85)	(3.31-6.12)
New Zealand	0.71	0.9	1.32	1.91	2.69	3.72	4.56
	(0.49-1.02)	(0.64-1.27)	(0.98-1.78)	(1.46-2.49)	(2.08-3.48)	(2.84-4.86)	(3.43-6.06)
Norway	0.46	0.61	0.96	1.54	2.46	3.82	5.07
	(0.31-0.69)	(0.42-0.90)	(0.67-1.36)	(1.09-2.17)	(1.73-3.49)	(2.62-5.58)	(3.40-7.57)
Portugal	0.28	0.42	0.78	1.47	2.71	4.73	6.68
	(0.15-0.51)	(0.24-0.74)	(0.47-1.30)	(0.93-2.32)	(1.74-4.21)	(2.99-7.48)	(4.13-10.80)
South Africa	0.71	0.9	1.32	1.99	3.01	4.37	5.48
	(0.50-1.00)	(0.65-1.25)	(0.97-1.79)	(1.48-2.67)	(2.24-4.04)	(3.21-5.94)	(3.98-7.55)
Spain	0.49	0.62	0.93	1.43	2.19	3.22	4.06
	(0.35-0.69)	(0.45-0.86)	(0.68-1.26)	(1.05-1.93)	(1.62-2.97)	(2.36-4.40)	(2.95-5.59)
Sweden	0.6	0.78	1.17	1.82	2.78	4	4.94
	(0.42-0.87)	(0.55-1.10)	(0.85-1.62)	(1.35-2.47)	(2.09-3.70)	(3.02-5.30)	(3.73-6.55)
Switzerland	0.59	0.74	1.06	1.58	2.33	3.29	4.03
	(0.43-0.81)	(0.55-1.00)	(0.80-1.42)	(1.20-2.08)	(1.78-3.06)	(2.51-4.32)	(3.06-5.32)
UK	0.62	0.8	1.18	1.75	2.55	3.55	4.35
	(0.43-0.90)	(0.56-1.13)	(0.87-1.61)	(1.32-2.32)	(1.95-3.32)	(2.71-4.66)	(3.28-5.77)
US	0.66	0.84	1.25	1.92	2.89	4.07	4.95
	(0.46-0.94)	(0.59-1.18)	(0.91-1.72)	(1.44-2.57)	(2.20-3.78)	(3.14-5.28)	(3.84-6.40)

Table A9: Countrywise estimates using the FF bootstrap estimator, with $T = 30$ years

This table shows the point estimates and 90 percent confidence intervals (in parentheses) for the 30-year distributions of country-specific gross returns, for each country with a full history in the DMS data set. The estimates are based on data for the entire sample period from 1900 to 2020, using the FF bootstrap estimator. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile).

	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
Australia	1.49	2.15	3.89	7.37	13.71	23.55	32.32
	(0.62-3.58)	(0.92-5.01)	(1.74-8.69)	(3.44-15.81)	(6.62-28.38)	(11.66-47.59)	(16.18-64.59)
Austria	0.09	0.17	0.46	1.36	3.88	9.89	17.29
	(0.02-0.39)	(0.04-0.70)	(0.12-1.77)	(0.37-4.95)	(1.09-13.79)	(2.80-34.88)	(4.88-61.22)
Belgium	0.29	0.46	1.00	2.33	5.37	11.29	17.51
	(0.10-0.87)	(0.16-1.35)	(0.35-2.85)	(0.84-6.48)	(1.97-14.66)	(4.17-30.58)	(6.47-47.41)
Canada	1.20	1.68	2.93	5.35	9.68	16.35	22.27
	(0.55-2.64)	(0.78-3.63)	(1.39-6.16)	(2.60-11.01)	(4.78-19.59)	(8.16-32.75)	(11.17-44.43)
Denmark	0.96	1.39	2.57	5.07	10.00	18.53	26.91
	(0.40-2.30)	(0.59-3.28)	(1.11-5.95)	(2.21-11.64)	(4.32-23.16)	(7.85-43.70)	(11.22-64.51)
Finland	0.42	0.74	1.85	5.04	13.45	32.10	54.01
	(0.11-1.61)	(0.20-2.71)	(0.53-6.45)	(1.51-16.83)	(4.13-43.79)	(9.93-103.81)	(16.65-175.21)
France	0.37	0.58	1.22	2.76	6.19	12.71	19.46
	(0.13-1.05)	(0.21-1.61)	(0.45-3.31)	(1.04-7.36)	(2.35-16.31)	(4.84-33.38)	(7.41-51.13)
Germany	0.10	0.24	0.92	3.13	9.18	23.22	40.22
	(0.01-1.04)	(0.03-1.99)	(0.15-5.49)	(0.71-13.86)	(2.57-32.76)	(7.04-76.62)	(12.00-134.85)
Ireland	0.43	0.71	1.59	3.73	8.44	17.25	26.32
	(0.12-1.47)	(0.22-2.31)	(0.53-4.78)	(1.34-10.38)	(3.21-22.17)	(6.78-43.87)	(10.44-66.33)
Italy	0.14	0.26	0.69	1.98	5.45	13.17	22.21
	(0.03-0.62)	(0.06-1.08)	(0.18-2.66)	(0.56-6.99)	(1.64-18.08)	(4.10-42.25)	(6.98-70.64)
Japan	0.17	0.36	1.17	3.88	11.39	28.23	47.70
	(0.02-1.21)	(0.06-2.27)	(0.23-5.97)	(0.95-15.91)	(3.28-39.54)	(8.89-89.62)	(15.34-148.38)
Netherlands	0.74	1.11	2.17	4.49	9.18	17.46	25.64
	(0.29-1.92)	(0.44-2.83)	(0.88-5.33)	(1.86-10.80)	(3.85-21.90)	(7.31-41.71)	(10.66-61.65)
New Zealand	1.25	1.84	3.42	6.62	12.52	22.17	31.20
	(0.49-3.21)	(0.74-4.54)	(1.47-7.98)	(2.97-14.73)	(5.77-27.17)	(10.23-48.01)	(14.29-68.14)
Norway	0.46	0.73	1.57	3.62	8.37	17.88	28.53
	(0.15-1.35)	(0.25-2.11)	(0.56-4.42)	(1.30-10.08)	(2.97-23.58)	(6.17-51.80)	(9.61-84.68)
Portugal	0.18	0.34	0.98	3.02	9.05	24.15	43.35
	(0.04-0.86)	(0.07-1.56)	(0.23-4.12)	(0.77-11.86)	(2.39-34.24)	(6.41-90.90)	(11.39-165.03)
South Africa	1.31	1.96	3.79	7.85	16.12	30.84	45.65
	(0.51-3.36)	(0.78-4.93)	(1.54-9.31)	(3.25-18.97)	(6.69-38.84)	(12.69-74.99)	(18.58-112.14)
Spain	0.46	0.69	1.36	2.89	6.09	11.91	17.72
	(0.18-1.19)	(0.27-1.76)	(0.54-3.42)	(1.16-7.17)	(2.46-15.10)	(4.77-29.69)	(7.06-44.50)
Sweden	0.89	1.37	2.74	5.82	12.20	23.36	34.25
	(0.33-2.44)	(0.51-3.64)	(1.07-7.02)	(2.36-14.38)	(5.08-29.28)	(9.91-55.04)	(14.64-80.11)
Switzerland	0.72	1.05	1.97	3.90	7.67	14.01	20.03
	(0.30-1.74)	(0.44-2.50)	(0.85-4.57)	(1.71-8.88)	(3.40-17.30)	(6.24-31.49)	(8.91-45.03)
UK	0.89	1.34	2.57	5.13	10.07	18.27	26.04
	(0.33-2.37)	(0.52-3.43)	(1.06-6.24)	(2.21-11.90)	(4.48-22.65)	(8.21-40.65)	(11.69-58.01)
US	1.10	1.66	3.26	6.78	13.79	25.61	36.65
	(0.41-2.93)	(0.64-4.32)	(1.31-8.11)	(2.84-16.19)	(5.99-31.77)	(11.42-57.42)	(16.56-81.10)

Table A10: Empirical estimates of global long-run returns, unbalanced panel with 32 countries

The table shows the pooled point estimates and 90 percent confidence intervals (in parentheses) for the long-run distributions of the global gross returns based on the complete unbalanced panel of 32 countries in the DMS data set. Panels A and B show results for 10-year and 30-year compounding horizons, respectively. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile). Results are presented using data for the entire sample period (Full sample) and for a subsample starting in 1960 (Post-1960), as indicated in the row headers. For each sample and horizon, results for the MLE, the skewness-corrected MLE (ML-skew), and the FF bootstrap estimator are presented.

Panel A: $T = 10$

	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
	I. ML						
Full sample	0.43	0.58	0.95	1.65	2.87	4.71	6.33
	(0.32-0.58)	(0.44-0.77)	(0.74-1.24)	(1.30-2.10)	(2.30-3.58)	(3.80-5.84)	(5.12-7.84)
Post-1960	0.43	0.59	1.01	1.82	3.30	5.62	7.73
	(0.26-0.71)	(0.37-0.96)	(0.65-1.57)	(1.22-2.73)	(2.27-4.80)	(3.93-8.04)	(5.44-11.00)
	II. ML-skew						
Full sample	0.41	0.57	0.97	1.72	2.93	4.60	5.95
	(0.30-0.55)	(0.43-0.75)	(0.75-1.26)	(1.35-2.18)	(2.34-3.66)	(3.71-5.70)	(4.81-7.36)
Post-1960	0.41	0.58	1.02	1.86	3.34	5.54	7.45
	(0.25-0.69)	(0.36-0.94)	(0.66-1.59)	(1.25-2.79)	(2.30-4.86)	(3.88-7.92)	(5.24-10.59)
	III. FF bootstrap						
Full sample	0.42	0.59	1.00	1.70	2.84	4.51	6.01
	(0.31-0.56)	(0.45-0.78)	(0.77-1.29)	(1.34-2.16)	(2.27-3.54)	(3.64-5.59)	(4.86-7.44)
Post-1960	0.42	0.59	1.04	1.87	3.28	5.45	7.45
	(0.25-0.69)	(0.37-0.96)	(0.67-1.62)	(1.25-2.79)	(2.25-4.77)	(3.81-7.79)	(5.24-10.59)

Panel B: $T = 30$

	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
	I. ML						
Full sample	0.44	0.74	1.75	4.53	11.75	27.72	46.33
	(0.20-0.99)	(0.34-1.62)	(0.83-3.68)	(2.22-9.24)	(5.93-23.29)	(14.27-53.85)	(24.07-89.17)
Post-1960	0.50	0.86	2.17	6.07	16.93	42.63	74.09
	(0.13-1.97)	(0.23-3.29)	(0.61-7.77)	(1.81-20.38)	(5.32-53.86)	(13.95-130.25)	(24.74-221.89)
	II. ML-skew						
Full sample	0.42	0.72	1.78	4.70	11.99	27.07	43.52
	(0.19-0.93)	(0.33-1.58)	(0.84-3.76)	(2.30-9.58)	(6.05-23.76)	(13.93-52.60)	(22.61-83.76)
Post-1960	0.48	0.85	2.20	6.20	17.13	42.03	71.33
	(0.12-1.90)	(0.22-3.24)	(0.62-7.87)	(1.85-20.83)	(5.39-54.51)	(13.75-128.41)	(23.82-213.63)
	III. FF bootstrap						
Full sample	0.42	0.74	1.81	4.69	11.78	26.76	43.68
	(0.19-0.94)	(0.34-1.61)	(0.86-3.82)	(2.30-9.57)	(5.94-23.34)	(13.77-51.99)	(22.69-84.08)
Post-1960	0.48	0.86	2.21	6.19	16.95	41.49	71.29
	(0.12-1.88)	(0.23-3.26)	(0.62-7.92)	(1.84-20.79)	(5.33-53.92)	(13.58-126.78)	(23.80-213.52)

Table A11: Empirical comparison of FF bootstrap and ACO block bootstrap estimates

The table shows the pooled point estimates and 90 percent confidence intervals (in parentheses) for the long-run distributions of the global gross returns based on the panel of 21 countries with a full history in the DMS data set. Panels A and B show results for 10-year and 30-year compounding horizons, respectively. The top row in each panel indicates what percentile of the distribution that is being considered (e.g., 10% indicates the 10^{th} percentile). Results are presented using data for the entire sample period (Full sample) and for a subsample starting in 1960 (Post-1960), as indicated in the row headers. For each sample and horizon, results for the FF bootstrap estimator the ACO block bootstrap estimator are presented. The block bootstrap estimator is implemented with a geometrically distributed random block size, with the average block size set to 10 years. Confidence intervals around the block bootstrap estimates are formed in an identical manner to those around the FF bootstrap (described in the main text).

	Panel A: $T = 10$						
	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
	I. FF bootstrap						
Full sample	0.46	0.63	1.00	1.62	2.56	3.84	4.93
	(0.34-0.61)	(0.48-0.83)	(0.78-1.29)	(1.29-2.04)	(2.06-3.17)	(3.13-4.72)	(4.02-6.05)
Post-1960	0.49	0.66	1.06	1.74	2.83	4.36	5.66
	(0.30-0.78)	(0.42-1.03)	(0.70-1.60)	(1.20-2.54)	(2.00-4.00)	(3.14-6.05)	(4.11-7.80)
	II. ACO block bootstrap						
Full sample	0.48	0.69	1.09	1.65	2.43	3.49	4.33
	(0.36-0.64)	(0.52-0.91)	(0.84-1.40)	(1.31-2.08)	(1.96-3.01)	(2.84-4.30)	(3.52-5.32)
Post-1960	0.62	0.79	1.16	1.76	2.57	3.65	4.49
	(0.39-0.98)	(0.51-1.22)	(0.78-1.72)	(1.22-2.52)	(1.84-3.58)	(2.66-5.01)	(3.29-6.11)

	Panel B: $T = 30$						
	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
	I. FF bootstrap						
Full sample	0.45	0.76	1.72	4.04	9.16	18.94	29.28
	(0.21-0.99)	(0.36-1.62)	(0.83-3.54)	(2.03-8.03)	(4.74-17.70)	(10.00-35.88)	(15.61-54.93)
Post-1960	0.58	0.95	2.13	5.11	12.06	25.87	40.80
	(0.16-2.09)	(0.27-3.31)	(0.65-6.99)	(1.66-15.78)	(4.13-35.22)	(9.23-72.48)	(14.87-111.95)
	II. ACO block bootstrap						
Full sample	0.57	0.95	1.99	4.15	8.17	14.94	21.63
	(0.26-1.25)	(0.44-2.04)	(0.96-4.14)	(2.07-8.33)	(4.20-15.92)	(7.83-28.50)	(11.45-40.87)
Post-1960	0.93	1.39	2.64	5.13	9.78	17.66	25.48
	(0.27-3.21)	(0.42-4.65)	(0.84-8.29)	(1.73-15.18)	(3.48-27.46)	(6.55-47.63)	(9.64-67.36)

Figure A1: Simulation results for $T = 120$ with $n = 720$ observations.

The figure shows simulation results based on estimates with a sample size of $n = 720$ and a compounding horizon of $T = 120$. Monthly period returns are generated according to the four different return models described in the main text: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals (Panel D). All four specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. Three different estimators are considered: (i) the MLE, (ii) the skewness-corrected MLE (ML-skew), and (iii) the FF bootstrap estimator. The results are based on $10,000$ samples. Each panel shows the median and the 5^{th} and 95^{th} percentiles of the estimated quantiles of the long-run gross return distributions (calculated across the $10,000$ estimates obtained from the simulated samples), for each of the three estimation procedures. The solid line shows the true (population) quantiles in each graph. The dotted line shows the median estimates for the MLE and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the ML estimates. The dashed lines show the median and the 5^{th} and 95^{th} percentiles of the skewness-corrected ML estimates of each quantile. The dashed-and-dotted lines show the corresponding estimates for the bootstrap estimator.

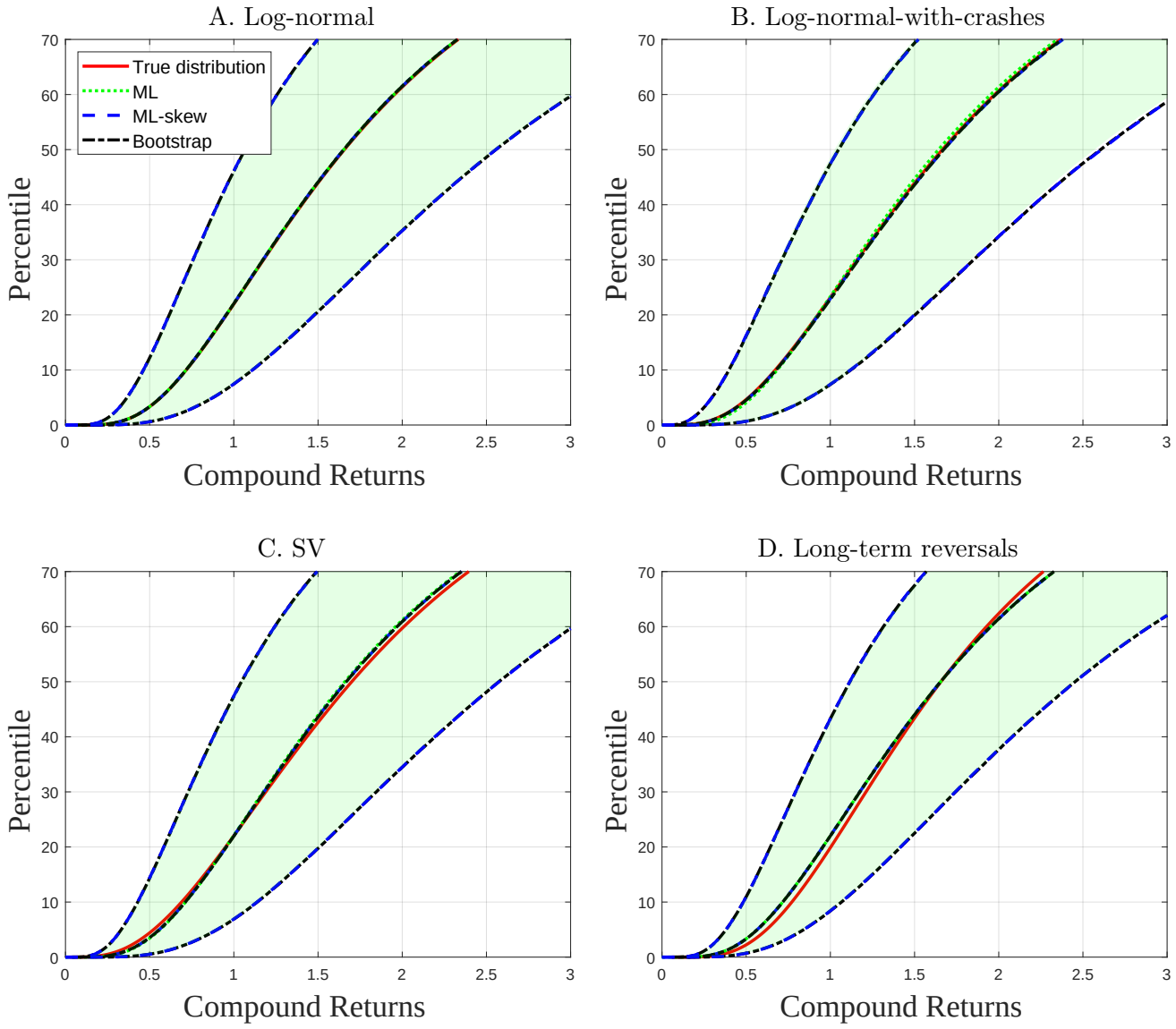


Figure A2: Simulation results for $T = 360$ with $n = 720$ observations.

The figure shows simulation results based on estimates with a sample size of $n = 720$ and a compounding horizon of $T = 360$. Monthly period returns are generated according to the four different return models described in the main text: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals (Panel D). All four specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. Three different estimators are considered: (i) the MLE, (ii) the skewness-corrected MLE (ML-skew), and (iii) the FF bootstrap estimator. The results are based on $10,000$ samples. Each panel shows the median and the 5^{th} and 95^{th} percentiles of the estimated quantiles of the long-run gross return distributions (calculated across the $10,000$ estimates obtained from the simulated samples), for each of the three estimation procedures. The solid line shows the true (population) quantiles in each graph. The dotted line shows the median estimates for the MLE and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the ML estimates. The dashed lines show the median and the 5^{th} and 95^{th} percentiles of the skewness-corrected ML estimates of each quantile. The dashed-and-dotted lines show the corresponding estimates for the bootstrap estimator.

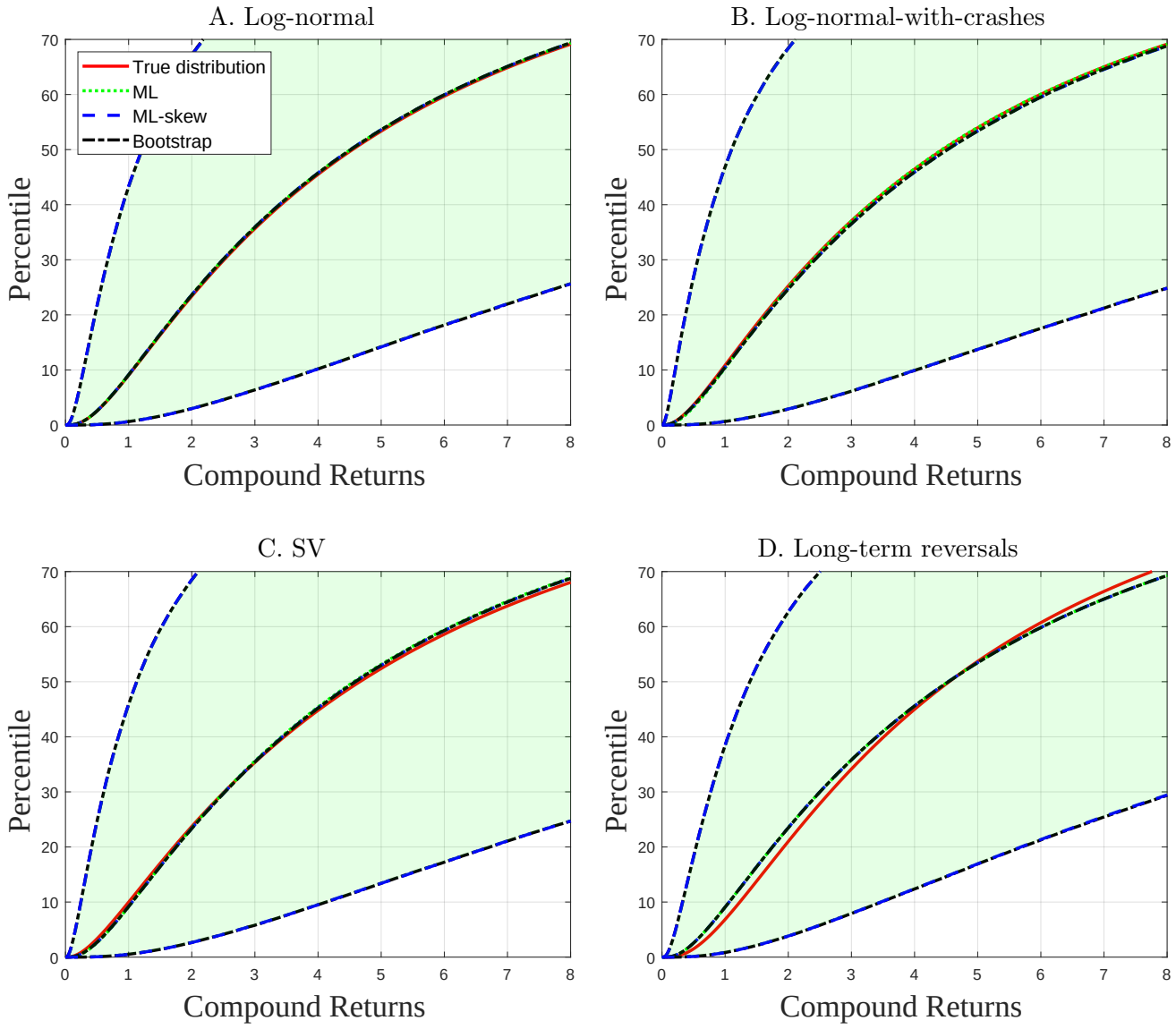
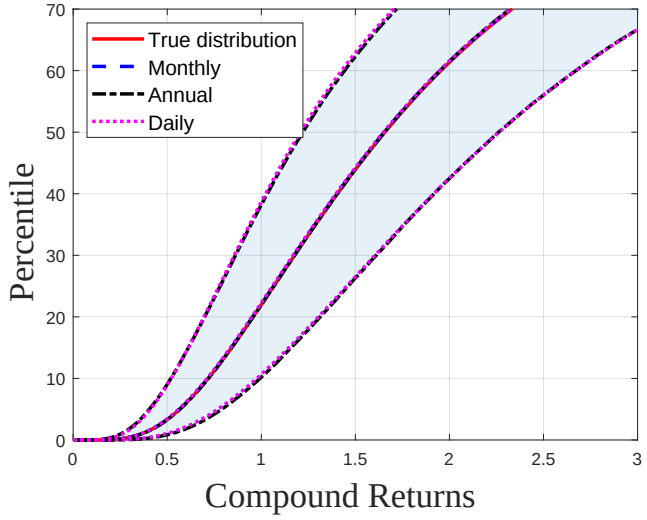


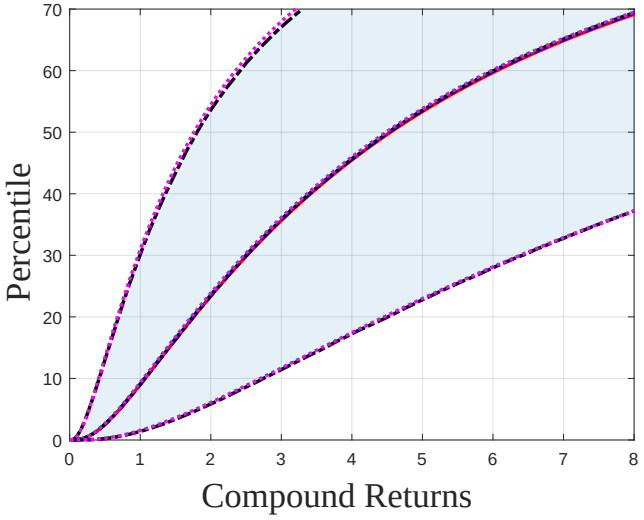
Figure A3: Simulation results with data sampled at different frequencies

The figure shows simulation results based on estimates with sample sizes corresponding to 120 years of data and compounding horizons of 10 years (Panels A1 and A2) and 30 years (Panels B1 and B2). Data are sampled either at a daily frequency ($n = 30,240$ observations), a monthly frequency ($n = 1,440$ observations), or an annual frequency ($n = 120$ observations). Two different return models are considered: i.i.d. log-normal (Panels A1 and B1) and stochastic volatility, SV (Panels A2 and B2). All specifications are parameterized such that the corresponding monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. The skewness-corrected MLE is used for estimation. The results are based on $10,000$ samples. Each panel shows the median and the 5^{th} and 95^{th} percentiles of the estimated quantiles of the long-run gross return distributions (calculated across the $10,000$ estimates obtained from the simulated samples), for each of the three sampling frequencies. The solid line shows the true (population) quantiles in each graph. The dashed line shows the median estimates for the estimates based on monthly data and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the monthly estimates. The dashed-and-dotted lines show the median and the 5^{th} and 95^{th} percentiles of the estimates based on annual data. The dotted lines show the corresponding estimates for the estimates based on daily data.

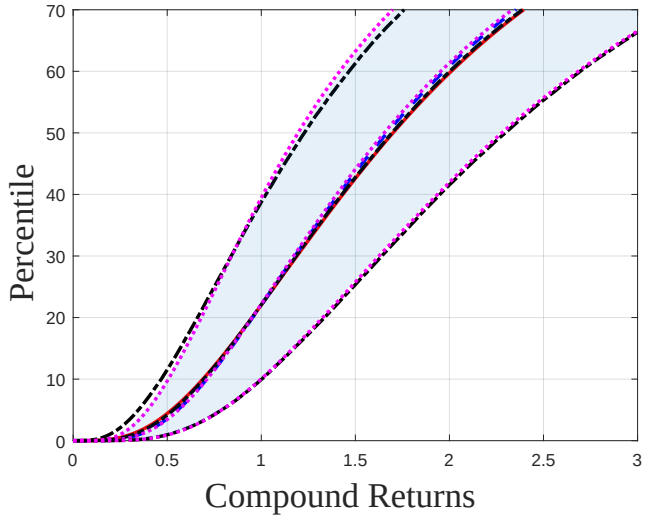
A1. Log-normal, 10-year compounding horizon



B1. Log-normal, 30-year compounding horizon



A2. SV, 10-year compounding horizon



B2. SV, 30-year compounding horizon

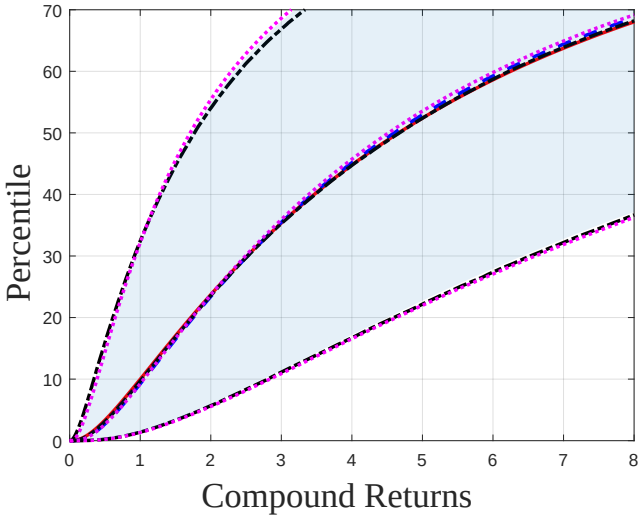


Figure A4: Simulation results for "direct" estimation

The figure shows simulation results based on estimates with sample sizes corresponding to 120 years of data and compounding horizons of 10 years (Panels A1 and A2) and 30 years (Panels B1 and B2). Data are observed at a monthly frequency ($n = 1,440$ observations). Estimates are either formed using the MLE based on the monthly data, labeled ML in the legend. Or using "direct" estimation, based on the available non-overlapping 10-year returns (12 return observation, Panels A1 and A2) or the available 30-year returns (4 return observations, Panels B1 and B2). The "direct" estimates of the long-run distributions are also formed using the normal MLE. Two different return models are considered: i.i.d. log-normal (Panels A1 and B1) and stochastic volatility, SV (Panels A2 and B2). All specifications are parameterized such that the corresponding monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. The results are based on $10,000$ samples. Each panel shows the median and the 5^{th} and 95^{th} percentiles of the estimated quantiles of the long-run gross return distributions (calculated across the $10,000$ estimates obtained from the simulated samples), for each of the two estimation approaches. The solid line shows the true (population) quantiles in each graph. The dotted line shows the median estimates for the ML estimates based on monthly data and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the monthly ML estimates. The dashed lines show the median and the 5^{th} and 95^{th} percentiles of the estimates based on the direct approach.

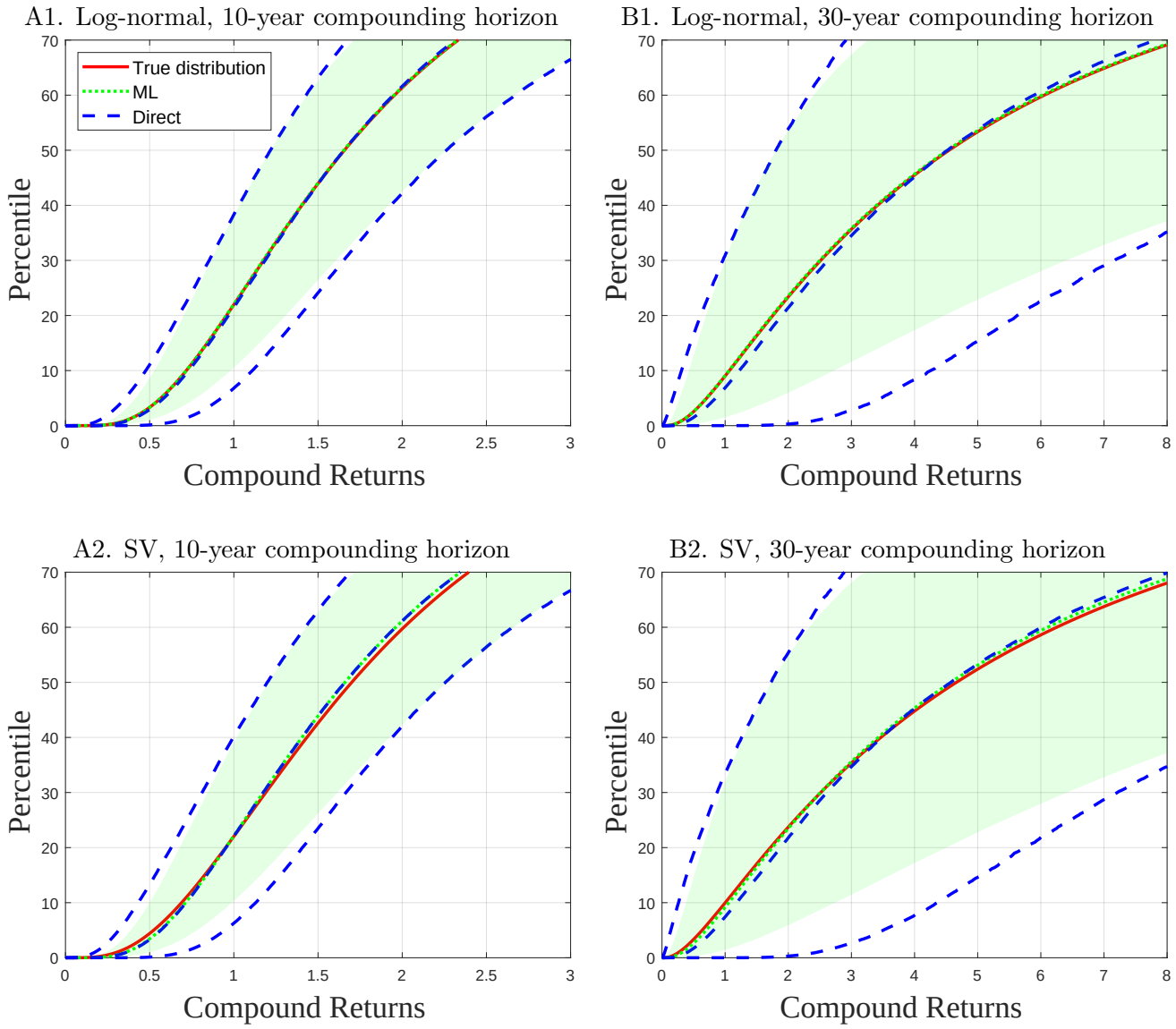


Figure A5: Simulation results with panel data

The figure shows simulation results based on panel-data estimates with a sample size of $n = 1,440$ and $K = 20$, and compounding horizons $T = 120$ (Panels A1 and A2) and $T = 360$ (Panels B1 and B2). Monthly period returns are generated according to a one-factor model with betas uniformly distributed between 0.7 and 1.3. Two different return models are considered: i.i.d. log-normal (Panels A1 and B1) and i.i.d. log-normal-with-crashes (Panels A2 and B2). All specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. The common factor in returns account for 40% of the total variance ($\lambda = 0.4$). Three different estimators are considered: (i) the pooled MLE, (ii) the pooled skewness-corrected MLE (ML-skew), and (iii) the pooled FF bootstrap estimator. The results are based on 10,000 samples. Each panel shows the median and the 5^{th} and 95^{th} percentiles of the estimated quantiles of the long-run gross return distributions (calculated across the 10,000 estimates obtained from the simulated samples), for each of the three estimation procedures. The solid line shows the true (population) quantiles in each graph, defined as the distribution for an asset with $\beta = 1$. The dotted line shows the median estimates for the MLE and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the ML estimates. The dashed lines show the median and the 5^{th} and 95^{th} percentiles of the skewness-corrected ML estimates of each quantile. The dashed-and-dotted lines show the corresponding estimates for the bootstrap estimator.

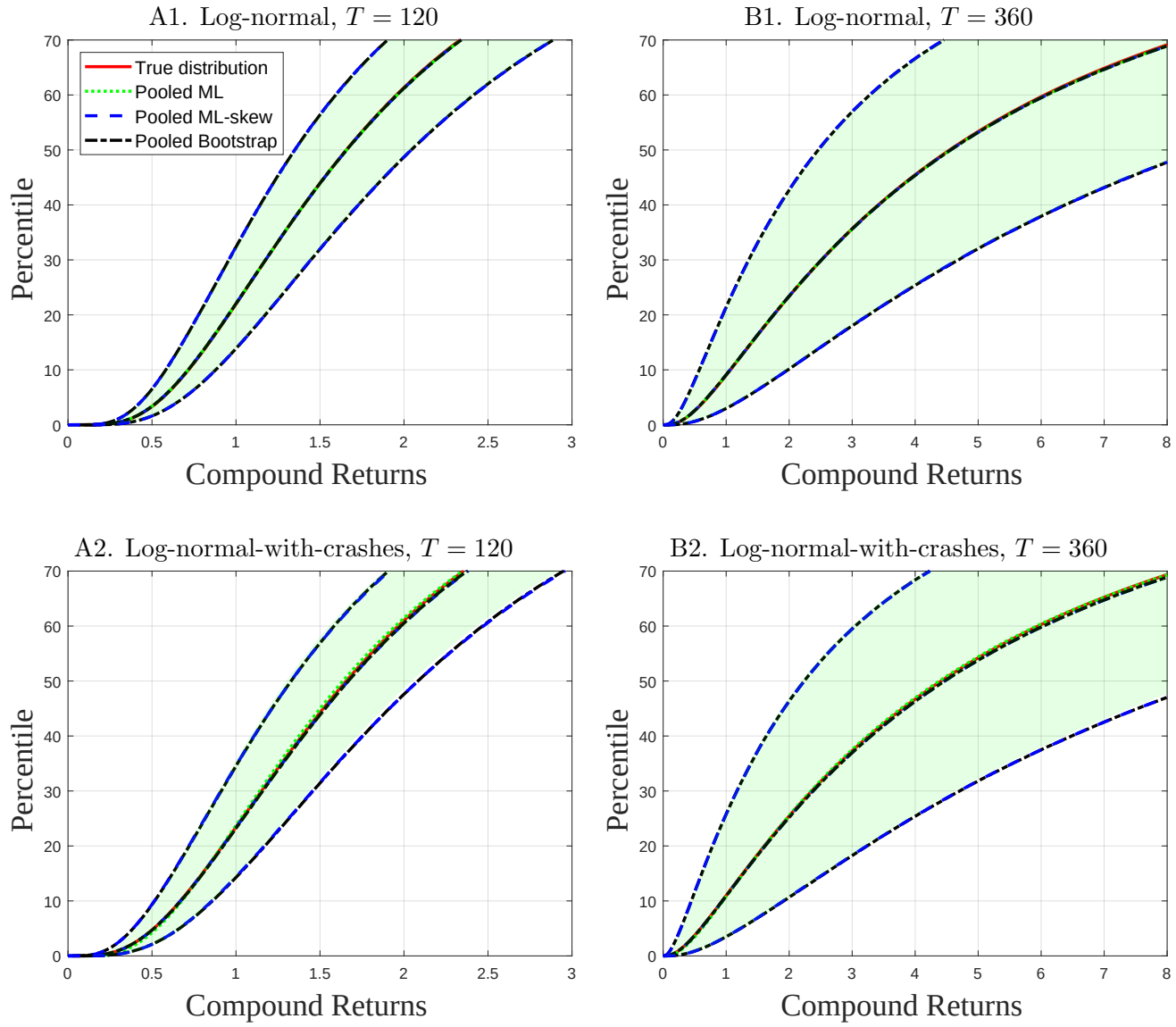


Figure A6: Simulation results for block bootstrap estimator with $T = 120$.

The figure shows simulation results based on estimates with a sample size of $n = 1,440$ and a compounding horizon of $T = 120$. Monthly period returns are generated according to five different return models described in the main text and the Online Appendix: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals with log returns following an MA(60) process (Panel D); long-term reversals with log returns following an MA(120) process (Panel E). All five specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. Three different estimators are considered: (i) the skewness-corrected MLE (ML-skew), (ii) the block bootstrap estimator with a random block length with a mean of 120 months (Block BS (120 rnd)), and (iii) the block bootstrap estimator with a fixed block length of 60 months (Block BS (60 fix)). The results are based on $10,000$ samples. Each panel shows the median and the 5^{th} and 95^{th} percentiles of the estimated quantiles of the long-run gross return distributions (calculated across the $10,000$ estimates obtained from the simulated samples), for each of the three estimation procedures. The solid line shows the true (population) quantiles in each graph. The dashed line shows the median estimates for the the skewness-corrected MLE and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the the skewness-corrected ML estimates. The dashed-and-dotted lines show the median and the 5^{th} and 95^{th} percentiles of the block bootstrap estimates (using a random block length with mean 120) of each quantile. The dotted lines show the corresponding estimates for the block bootstrap estimator with a fixed block length of 60.

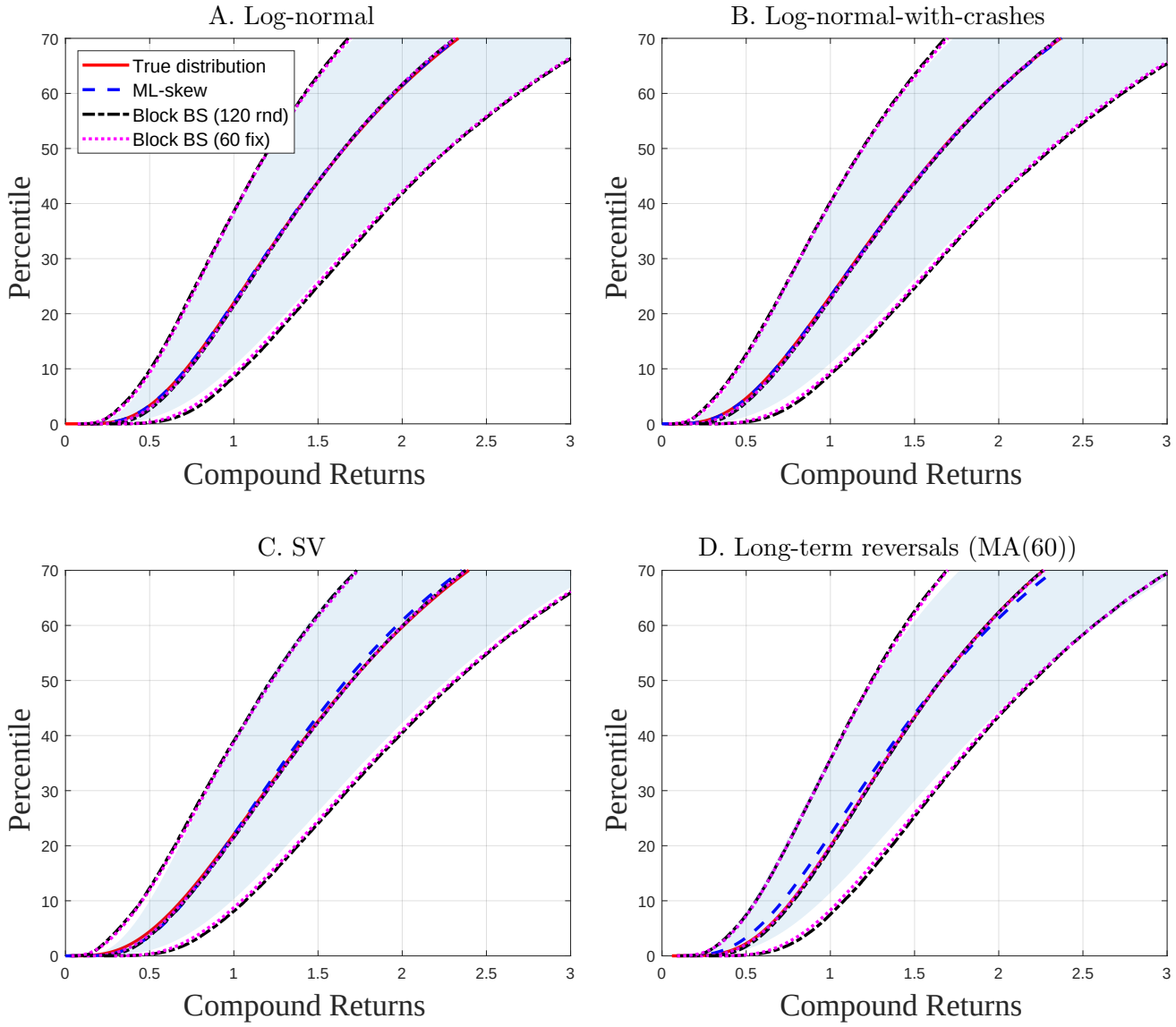


Figure A6: Simulation results for block bootstrap estimator with $T = 120$ (continued).

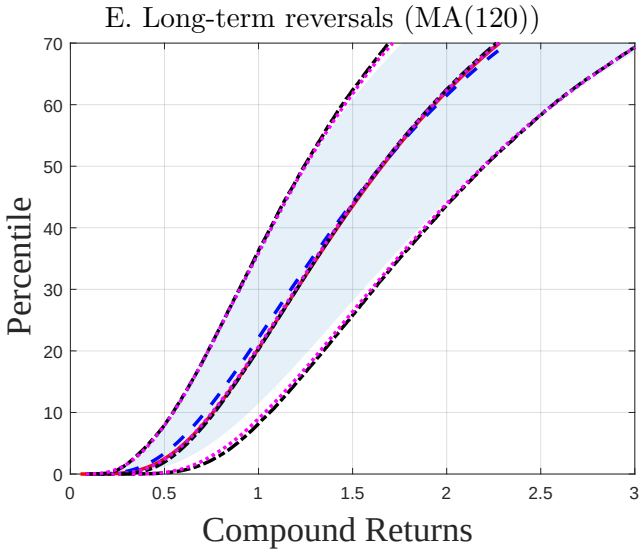


Figure A7: Simulation results for block bootstrap estimator with $T = 360$.

The figure shows simulation results based on estimates with a sample size of $n = 1,440$ and a compounding horizon of $T = 360$. Monthly period returns are generated according to five different return models described in the main text and the Online Appendix: i.i.d. log-normal (Panel A); i.i.d. log-normal-with-crashes (Panel B); stochastic volatility, SV (Panel C); long-term reversals with log returns following an MA(60) process (Panel D); long-term reversals with log returns following an MA(120) process (Panel E). All five specifications are parameterized such that monthly gross returns have a mean $\mu = 1.006$ and a volatility $\sigma = 0.06$. Three different estimators are considered: (i) the skewness-corrected MLE (ML-skew), (ii) the block bootstrap estimator with a random block length with a mean of 120 months (Block BS (120 rnd)), and (iii) the block bootstrap estimator with a fixed block length of 60 months (Block BS (60 fix)). The results are based on 10,000 samples. Each panel shows the median and the 5^{th} and 95^{th} percentiles of the estimated quantiles of the long-run gross return distributions (calculated across the 10,000 estimates obtained from the simulated samples), for each of the three estimation procedures. The solid line shows the true (population) quantiles in each graph. The dashed line shows the median estimates for the the skewness-corrected MLE and the edges of the shaded region corresponds to the 5^{th} and 95^{th} percentiles of the the skewness-corrected ML estimates. The dashed-and-dotted lines show the median and the 5^{th} and 95^{th} percentiles of the block bootstrap estimates (using a random block length with mean 120) of each quantile. The dotted lines show the corresponding estimates for the block bootstrap estimator with a fixed block length of 60.

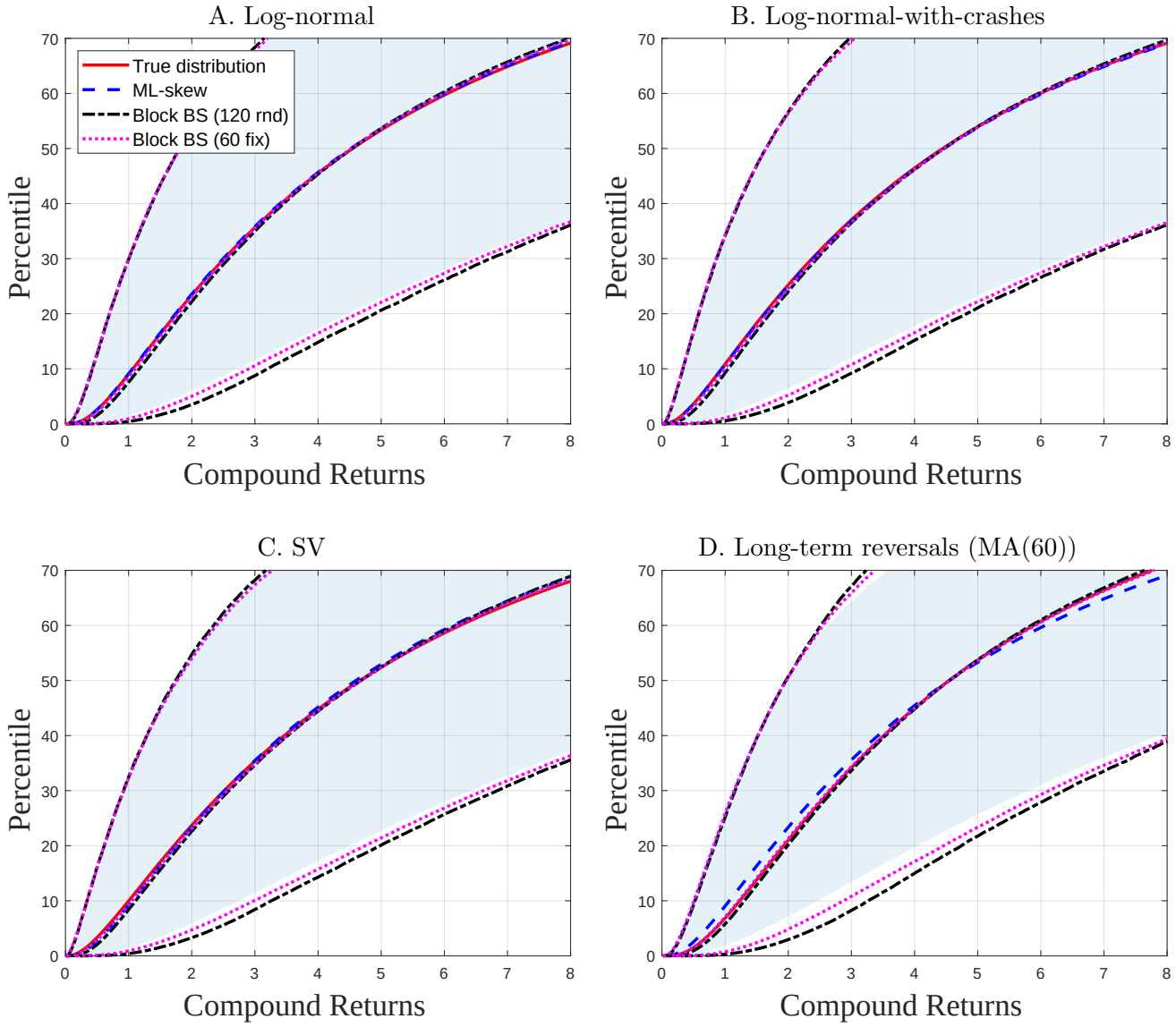


Figure A7: Simulation results for block bootstrap estimator with $T = 360$ (continued).

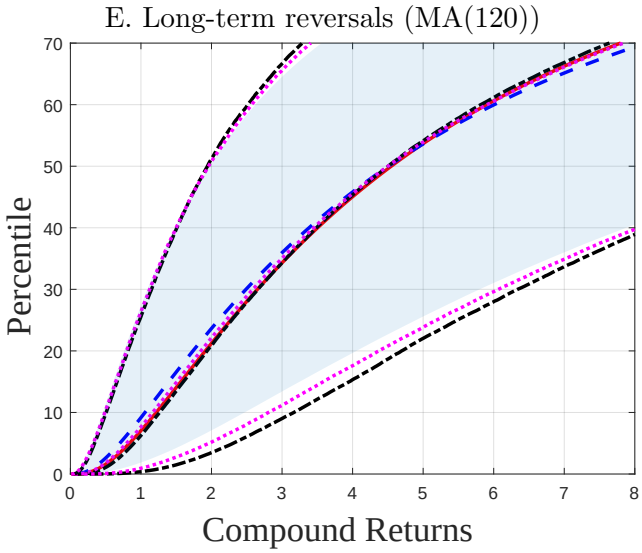


Figure A8: Data availability in the DMS data set

The figure shows data availability for each country in the full DMS data set. The start of each bar indicate the starting year of data availability for a given country.

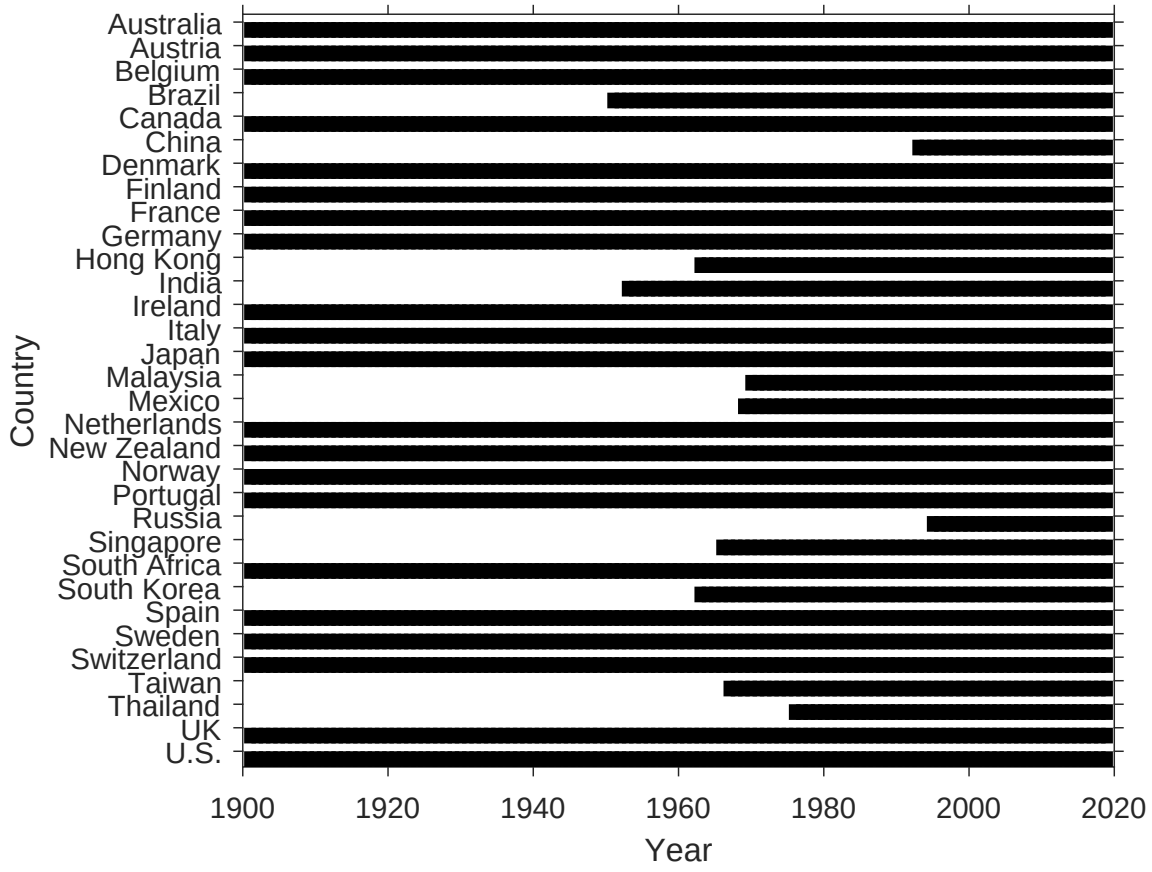


Figure A9: Individual country returns versus global returns, $T = 10$ years

The figure shows estimates of the long-run distribution of global 10-year gross returns along with the corresponding estimates for individual countries, as indicted in each panel header. The results are based on data for the entire sample period. All estimates are based on the skewness-corrected MLE. The estimates for the global distribution are formed from the pooled panel of 21 countries with a full history in the DMS data set. The dashed line and the shaded area show point estimates and 90% confidence bands, respectively, for the global return distributions. The solid line shows the point estimates for the long-run return distribution based on returns data for the individual country indicated in each panel header; the dotted lines show the corresponding 90% confidence bands. In addition, p-values are shown for the test of the null hypothesis that a given percentile of the global return distribution is identical to the corresponding percentile for the country-specific distribution. Specifically, p-values for the 5^{th}, 50^{th} and 95^{th} percentiles, labeled p-val(p_5), p-val(p_{50}), and p-val(p_{95}), respectively, are displayed.

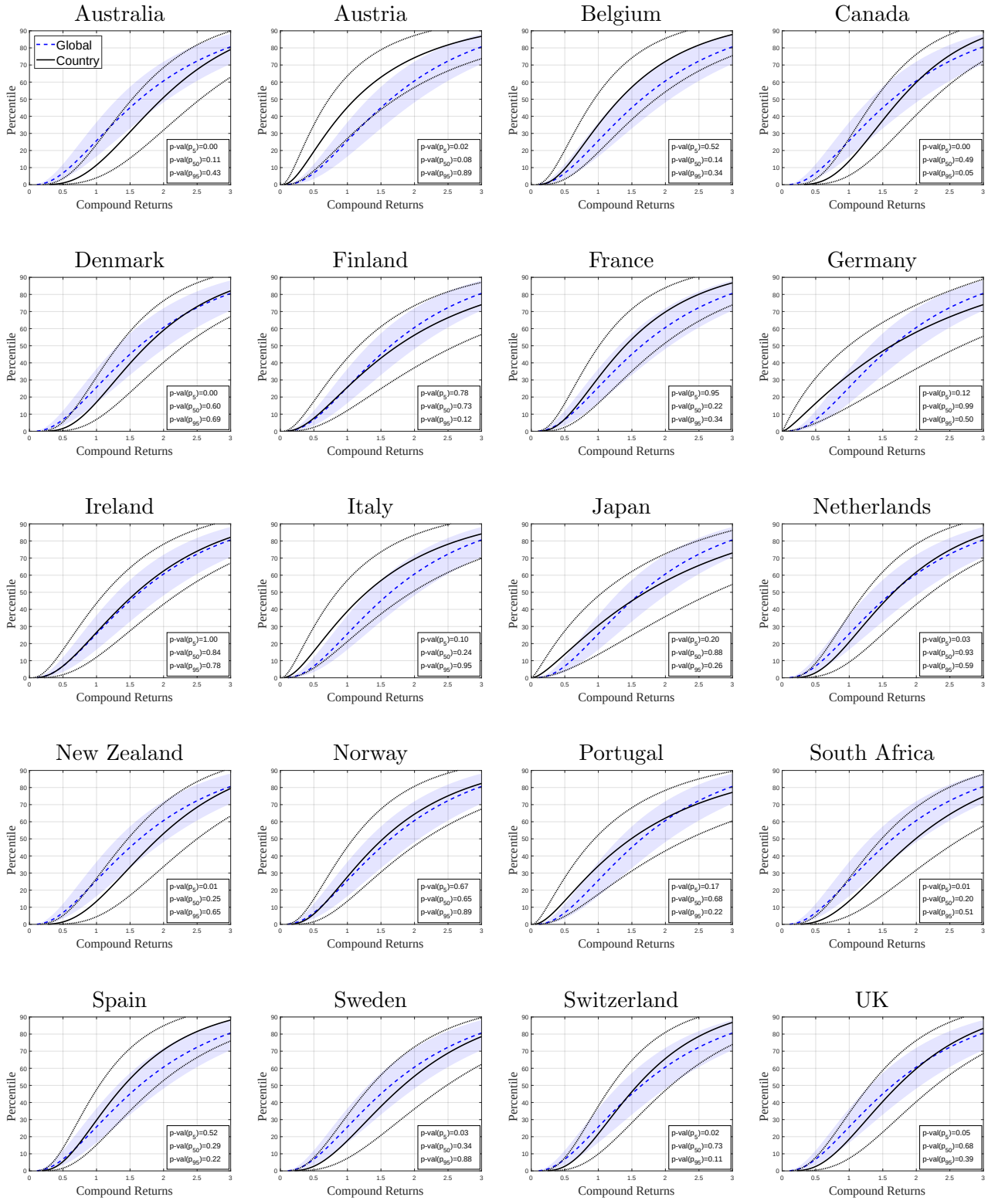


Figure A10: Individual country returns versus global returns, $T = 30$ years

The figure shows estimates of the long-run distribution of global 30-year gross returns along with the corresponding estimates for individual countries, as indicted in each panel header. The results are based on data for the entire sample period. All estimates are based on the skewness-corrected MLE. The estimates for the global distribution are formed from the pooled panel of 21 countries with a full history in the DMS data set. The dashed line and the shaded area show point estimates and 90% confidence bands, respectively, for the global return distributions. The solid line shows the point estimates for the long-run return distribution based on returns data for the individual country indicated in each panel header; the dotted lines show the corresponding 90% confidence bands. In addition, p-values are shown for the test of the null hypothesis that a given percentile of the global return distribution is identical to the corresponding percentile for the country-specific distribution. Specifically, p-values for the 5^{th}, 50^{th} and 95^{th} percentiles, labeled p-val(p_5), p-val(p_{50}), and p-val(p_{95}), respectively, are displayed.