Statistica Sinica Preprint No: SS-2022-0371	
Title	Improved Regression Inference Using a Second Overlapping Regression Model
Manuscript ID	SS-2022-0371
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0371
Complete List of Authors	Liang Peng and John Einmahl
Corresponding Authors	John Einmahl
E-mails	j.h.j.einmahl@uvt.nl

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

Improved regression inference using a second overlapping regression model

Liang Peng[1] and John H.J. Einmahl[2]

Abstract. Two time series of financial losses may be observed in different overlapping windows, serially dependent, heteroscedastic, and cross-sectionally dependent. Fitting a regression model to each of the two time series, we construct an improved least squares estimator in one series exploiting the cross-sectional dependence with the other series. We employ a random weight bootstrap method to define the new estimator and to establish its asymptotic normality. The developed inference is robust against heteroscedasticity as we do not estimate the heteroscedastic errors. Simulations confirm the efficiency improvement through substantial variance reduction, especially when the cross-sectional dependence is strong and the second series is longer. We illustrate the usefulness of the method by analyzing mutual funds' returns.

Key words and phrases: Cross-sectional dependence; Heteroscedasticity; Random weight bootstrap; Regression model; Variance reduction.

1 Introduction

Regression is a standard technique in exploring the relationship between predictors and response; see, e.g., Davison (2011). Consider two parametric regressions for forecasting financial returns

[1]Department of Risk Management and Insurance, Georgia State University, USA

[2]Dept. of Econometrics and OR, Tilburg University, The Netherlands; corresponding author: j.h.j.einmahl@uvt.nl

or losses $\{Y_{t,k}\}_{t=1}^T$ of two different assets or institutions ($k = 1, 2$) using predictors $\{\boldsymbol{X}_{t,k}\}_{t=1}^T$:

$$Y_{t,k} = f_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k) + \varepsilon_{t,k}, \tag{1}$$

where $E(\varepsilon_{t,k}) = 0$ and f_k has a known parametric form with unknown parameters $\boldsymbol{\theta}_k \in \mathbb{R}^{d_k}$ for $k = 1$ and 2. For example, in evaluating mutual funds' performances, researchers and practitioners often use the one-factor model in Jensen (1968), or the three-factor model in Fama and French (1996), or the four-factor model in Carhart (1997), where $\boldsymbol{X}_{t,1} = \boldsymbol{X}_{t,2}$. A simple estimator of $\boldsymbol{\theta}_k$ without modeling the distribution of the $\varepsilon_{t,k}$ is the least squares estimator

$$\widetilde{\boldsymbol{\theta}}_k = \operatorname*{argmin}_{\boldsymbol{\theta}_k} \sum_{t=1}^T \{Y_{t,k} - f_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k)\}^2.$$

When $\{(\varepsilon_{t,1}, \varepsilon_{t,2})\}_{t=1}^T$ are independent and identically distributed random vectors with a parametric bivariate density, joint inference (that is, taking into account the dependence structure) leads to improved estimators compared to the "marginal" least squares estimators $\widetilde{\boldsymbol{\theta}}_1$ and $\widetilde{\boldsymbol{\theta}}_2$, see Patton (2006), which uses a parametric copula model and allows for a different length of these two sequences. When $\{(\varepsilon_{t,1}, \varepsilon_{t,2})\}_{t=1}^T$ follows from a bivariate GARCH model, one can model the residuals parametrically to get the conditional likelihood for $(\varepsilon_{t,1}, \varepsilon_{t,2})$ and employ an efficient likelihood inference. However, if we do not model and infer the heteroscedasticity, it is infeasible to specify a likelihood or conditional likelihood for $(\varepsilon_{t,1}, \varepsilon_{t,2})$ and employ a parametric likelihood procedure. Hence, a natural question is whether one can improve the efficiency of $\widetilde{\boldsymbol{\theta}}_1$ by using the series $\{Y_{t,2}, \boldsymbol{X}_{t,2}\}$ nonparametrically when $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$ in (1) are correlated.

When one has a sequence of responses and a longer sequence of predictors, it is shown in Zhang, Brown and Cai (2019) that the extra predictors can improve the inference for a regression model based on pairs of responses and predictors and it is called a semi-supervised model. When one has two dependent sequences with different lengths, Ahmed and Einmahl (2019) show that extreme-value index estimation for the short sequence can be improved by using the longer

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

sequence. As far as we know, the above question for two regression models has not been studied in the literature, although such a data structure often appears in, e.g., financial econometrics.

This paper answers this question positively when both series are observed in different time windows, the second series has a time overlap with the first series, and there is dependence between the regression errors of the two series. More specifically, we construct a class of estimators combining the least squares estimator calculated from the first series and the difference of some statistics based on the second series for the overlapping period and the longer entire second period, respectively. Because the expectation of the product of the centered least squares estimator for the first series and the above difference is unequal to 0, we can minimize the asymptotic variance of this class of estimators to obtain a novel estimator, more efficient than the least squares estimator. To actually obtain this optimal estimator, we employ a random weight bootstrap method to estimate the asymptotic variances of this class of estimators. Because we allow heteroscedastic errors and do not infer them, the proposed inference is robust against heteroscedasticity misspecification, and the usual residual-based bootstrap method does not apply.

A particular application of the improved inference is mutual funds performance evaluation, where funds are often observed in different but overlapping time windows and have cross-sectional dependence. After fitting the popular one-factor, three-factor, or four-factor model to excess returns of the fund, researchers and practitioners use the intercept and the coefficient related to market excess returns to measure the skill and risk attitude of the fund manager, respectively. Because many funds have a smaller sample size, an interesting question is how to improve the classical least squares inference when there are other funds with a longer time series. Using US mutual funds' daily returns from September 1, 1998 to December 31, 2018, we compare least squares estimation and our novel inference for some funds with a shorter length

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

and use as a second series the average returns of 654 funds with the largest sample size in our data set.

We organize this paper as follows. Section 2 presents the methodology and asymptotic results. Sections 3 and 4 contain the simulation study and the mutual fund's data analysis, respectively. Section 5 concludes. The proofs of the main results are deferred to the Appendix.

2 Methodology and Asymptotic Results

Suppose we observe $\{Y_{t,1}, \boldsymbol{X}_{t,1}\}_{t=t_1}^{T_1}$ and $\{Y_{t,2}, \boldsymbol{X}_{t,2}\}_{t=t_2}^{T_2}$, with $t_2 < T_1 < T_2$, from the regression models (1), that is, there is an overlapping period between these two series. We allow both cases: $t_1 \geq t_2$ and $t_1 < t_2$. The first one is generally called the semi-supervised model; see, e.g., Zhang, Brown and Cai (2019) and the references therein. The least squares estimators for $\boldsymbol{\theta}_1 \in \mathbb{R}^{d_1}$ and $\boldsymbol{\theta}_2 \in \mathbb{R}^{d_2}$ are

$$\widetilde{\boldsymbol{\theta}}_1 = \underset{\boldsymbol{\theta}_1}{\operatorname{argmin}} \sum_{t=t_1}^{T_1} \{Y_{t,1} - f_1(\boldsymbol{X}_{t,1}; \boldsymbol{\theta}_1)\}^2 \text{ and } \widetilde{\boldsymbol{\theta}}_2 = \underset{\boldsymbol{\theta}_2}{\operatorname{argmin}} \sum_{t=t_2}^{T_2} \{Y_{t,2} - f_2(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2)\}^2.$$

The question is if one can improve the efficiency of $\widetilde{\boldsymbol{\theta}}_1$ using the data $\{Y_{t,2}, \boldsymbol{X}_{t,2}\}_{t=t_2}^{T_2}$ nonparametrically. To better appreciate the proposed methodology, we consider estimating $\gamma_1 = K(\boldsymbol{\theta}_1)$, where K is a known function from $\mathbb{R}^{d_1}$ to $\mathbb{R}$, and we would like to improve the least squares estimator $\widetilde{\gamma}_1 = K(\widetilde{\boldsymbol{\theta}}_1)$. For example, when f_1 is a linear function, $K(\boldsymbol{\theta}_1)$ being the intercept (coefficient) is an important quantity in studying the skill (risk attitude) of a fund manager in finance. The critical idea of improving the least squares estimation is to make use of the dependence between $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$ in combination with the additional information on $\varepsilon_{t,2}$ outside the overlapping period, without modeling and estimating the dependence structure explicitly. Generalizing the idea to a multivariate γ_1 is straightforward; see Remark 2 below.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

2.1 Uncorrelated errors

We first consider uncorrelated but heteroscedastic errors as in condition C1) below. Put $\widetilde{\varepsilon}_{t,k} = Y_{t,k} - f_k(\boldsymbol{X}_{t,k}; \widetilde{\boldsymbol{\theta}}_k)$, $k = 1, 2$. To improve the efficiency of $\widetilde{\gamma}_1$, we study a class of linear combinations of $\widetilde{\gamma}_1$ and some estimators with no asymptotic bias but correlated with $\widetilde{\gamma}_1$. Specifically, we consider the class of estimators

$$\widetilde{\gamma}_1 + h^\tau \Delta_2, \tag{2}$$

where h^τ denotes the transpose of the vector h, and

$$\Delta_2 = \frac{1}{T_1 - (t_1 \vee t_2) + 1} \sum_{t=t_1 \vee t_2}^{T_1} G(\widetilde{\varepsilon}_{t,2}, \boldsymbol{X}_{t,2}, \widetilde{\boldsymbol{\theta}}_2) - \frac{1}{T_2 - t_2 + 1} \sum_{t=t_2}^{T_2} G(\widetilde{\varepsilon}_{t,2}, \boldsymbol{X}_{t,2}, \widetilde{\boldsymbol{\theta}}_2) \tag{3}$$

with G being a known vector of functions and $t_1 \vee t_2$ denoting the maximum of t_1 and t_2. The choice of this class of estimators with Δ_2 and G is a very general form of the usual estimators in semi-supervised models, see Zhang, Brown and Cai (2019). The vector Δ_2 compares the behavior of the second series on the entire period with that on the overlapping period and has an asymptotically zero mean. Below, in (5), we will choose a specific G based on score functions. Under appropriate regularity conditions given later,

$$\begin{aligned}
\Delta_2 &= \frac{1}{T_1 - (t_1 \vee t_2) + 1} \sum_{t_1 \vee t_2}^{T_1} G(\varepsilon_{t,2}, \boldsymbol{X}_{t,2}, \boldsymbol{\theta}_2) \\
&+ \frac{1}{T_1 - (t_1 \vee t_2) + 1} \sum_{t_1 \vee t_2}^{T_1} \left\{ \frac{\partial G(\varepsilon_{t,2}, \boldsymbol{X}_{t,2}, \boldsymbol{\theta}_2)}{\partial \varepsilon_{t,2}} (\widetilde{\varepsilon}_{t,2} - \varepsilon_{t,2}) + \frac{\partial G(\varepsilon_{t,2}, \boldsymbol{X}_{t,2}, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2^\tau} (\widetilde{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) \right\} \\
&\quad + o_p\left(\frac{1}{\sqrt{T_1 - (t_1 \vee t_2)}} \right) - \frac{1}{T_2 - t_2 + 1} \sum_{t_2}^{T_2} G(\varepsilon_{t,2}, \boldsymbol{X}_{t,2}, \boldsymbol{\theta}_2) \\
&- \frac{1}{T_2 - t_2 + 1} \sum_{t_2}^{T_2} \left\{ \frac{\partial G(\varepsilon_{t,2}, \boldsymbol{X}_{t,2}, \boldsymbol{\theta}_2)}{\partial \varepsilon_{t,2}} (\widetilde{\varepsilon}_{t,2} - \varepsilon_{t,2}) + \frac{\partial G(\varepsilon_{t,2}, \boldsymbol{X}_{t,2}, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2^\tau} (\widetilde{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) \right\} + o_p\left(\frac{1}{\sqrt{T_1 - (t_1 \vee t_2)}} \right) \\
&= \frac{1}{T_1 - (t_1 \vee t_2) + 1} \sum_{t_1 \vee t_2}^{T_1} G(\varepsilon_{t,2}, \boldsymbol{X}_{t,2}, \boldsymbol{\theta}_2) - \frac{1}{T_2 - t_2 + 1} \sum_{t_2}^{T_2} G(\varepsilon_{t,2}, \boldsymbol{X}_{t,2}, \boldsymbol{\theta}_2) + o_p\left(\frac{1}{\sqrt{T_1 - (t_1 \vee t_2)}} \right), \quad (4)
\end{aligned}$$

where the second and fifth terms are combined for the last equality. This result simplifies the study of the asymptotic behavior of the new estimators in (2), since the plug-in estimator $\widetilde{\boldsymbol{\theta}}_2$ in Δ_2 plays, directly and indirectly through the $\widetilde{\varepsilon}_{t,k}$, no role asymptotically.

By minimizing $E(\widetilde{\gamma}_1 + h^\tau \Delta_2 - \gamma_1)^2$ with respect to h, we obtain the optimal h as

$$h_0 = -\{E(\Delta_2 \Delta_2^\tau)\}^{-1} E\{(\widetilde{\gamma}_1 - \gamma_1)\Delta_2\}.$$

Note that Δ_2 is a difference of two statistics computed from the overlapping period and the entire period of the second series, respectively. When the proportion $\{T_1 - (t_1 \vee t_2)\}/(T_1 - t_1)$, the relative overlap of the series $\{Y_{t,2}, \boldsymbol{X}_{t,2}\}_{t=t_2}^{T_2}$ with the series $\{Y_{t,1}, \boldsymbol{X}_{t,1}\}_{t=t_1}^{T_1}$, stays positive, the expectation $E\{(\widetilde{\gamma}_1 - \gamma_1)\Delta_2\}$ will in general be nonzero and the newly obtained "estimator" with h_0 will have a smaller asymptotic variance than $\widetilde{\gamma}_1$ itself. More precisely, the variance gain is equal to

$$E(\widetilde{\gamma}_1 - \gamma_1)^2 - E(\widetilde{\gamma}_1 + h_0^\tau \Delta_2 - \gamma_1)^2 = E\{(\widetilde{\gamma}_1 - \gamma_1)\Delta_2^\tau\} \{E(\Delta_2 \Delta_2^\tau)\}^{-1} E\{(\widetilde{\gamma}_1 - \gamma_1)\Delta_2\},$$

which is always nonnegative. Clearly, when the residuals in both regressions are independent, $E\{(\widetilde{\gamma}_1 - \gamma_1)\Delta_2\} = E(\widetilde{\gamma}_1 - \gamma_1)E(\Delta_2) \to \mathbf{0}$ and $h_0 \to \mathbf{0}$, i.e., no improvement. There are many possible choices for G when constructing Δ_2, and it is challenging to find an optimal one. However, the very natural choice for G that we will employ here is the one using score functions, that is,

$$G(\widetilde{\varepsilon}_{t,2}, \boldsymbol{X}_{t,2}, \widetilde{\boldsymbol{\theta}}_2) = \widetilde{\varepsilon}_{t,2} \dot{f}_2(\boldsymbol{X}_{t,2}; \widetilde{\boldsymbol{\theta}}_2), \tag{5}$$

where $\dot{f}_2(x; \boldsymbol{\theta}_2) = \frac{\partial f_2(x;\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2}$. Observe that in this case,

$$\begin{aligned} \Delta_2 &= \frac{1}{T_1 - (t_1 \vee t_2) + 1} \sum_{t=t_1 \vee t_2}^{T_1} \widetilde{\varepsilon}_{t,2} \dot{f}_2(\boldsymbol{X}_{t,2}; \widetilde{\boldsymbol{\theta}}_2) \\ &= \frac{1}{T_1 - (t_1 \vee t_2) + 1} \sum_{t=t_1 \vee t_2}^{T_1} (Y_{t,2} - f_2(\boldsymbol{X}_{t,2}; \widetilde{\boldsymbol{\theta}}_2)) \dot{f}_2(\boldsymbol{X}_{t,2}; \widetilde{\boldsymbol{\theta}}_2), \end{aligned}$$

since the second part of Δ_2 in (3) is zero because $\widetilde{\boldsymbol{\theta}}_2$ is the least squares estimator.

To estimate h_0 without deriving the approximations of $E\{(\widetilde{\gamma}_1 - \gamma_1)\Delta_2\}$ and $E(\Delta_2 \Delta_2^\tau)$, we adopt the random weight bootstrap method in Jin, Ying and Wei (2001) and Zhu (2016). An interesting feature of our approach is that the proposed estimator is robust against the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

heteroscedasticity of the $\varepsilon_{t,k}$ as we do not infer it. Also, because we do not estimate this heteroscedasticity, the usual residual-based bootstrap method does not apply. Although one can employ other bootstrap methods, such as the wild bootstrap in Mammen (1993) without estimating heteroscedasticity, it generally requires generating bootstrap samples from the model and becomes more computationally intensive than the random weighted bootstrap method. The random weight bootstrap procedure is as follows:

- Step Ai) Draw a random sample with size T_2 from a probability distribution with mean 1 and variance 1, say the standard exponential distribution. Denote it by $\{\delta_t^b\}_{t=1}^{T_2}$.
- Step Aii) Compute

$$
\begin{aligned}
\widetilde{\boldsymbol{\theta}}_1^b &= \operatorname*{argmin}_{\boldsymbol{\theta}_1} \sum_{t=t_1}^{T_1} \delta_t^b \{Y_{t,1} - f_1(\boldsymbol{X}_{t,1}; \boldsymbol{\theta}_1)\}^2, \quad \widetilde{\gamma}_1^b = K(\widetilde{\boldsymbol{\theta}}_1^b),\\
\widetilde{\boldsymbol{\theta}}_2^b &= \operatorname*{argmin}_{\boldsymbol{\theta}_2} \sum_{t=t_2}^{T_2} \delta_t^b \{Y_{t,2} - f_2(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2)\}^2,\\
\widetilde{\varepsilon}_{t,2}^b &= Y_{t,2} - f_2(\boldsymbol{X}_{t,2}; \widetilde{\boldsymbol{\theta}}_2^b), \text{ for } t = t_2, \ldots, T_2,\\
\Delta_2^b &= \frac{\sum_{t=t_1 \vee t_2}^{T_1} \delta_t^b \, \widetilde{\varepsilon}_{t,2}^b \dot{f}_2(\boldsymbol{X}_{t,2}; \widetilde{\boldsymbol{\theta}}_2^b)}{\sum_{t=t_1 \vee t_2}^{T_1} \delta_t^b}.
\end{aligned}
$$

- Step Aiii) Repeat the above two steps B times to obtain $\{\widetilde{\gamma}_1^b, \Delta_2^b\}_{b=1}^B$.

Next, we estimate $E\{(\widetilde{\gamma}_1 - \gamma_1)\Delta_2\}$ and $E(\Delta_2 \Delta_2^\tau)$ by

$$
\frac{1}{B}\sum_{b=1}^{B} (\widetilde{\gamma}_1^b - \widetilde{\gamma}_1)(\Delta_2^b - \Delta_2) \quad \text{and} \quad \frac{1}{B}\sum_{b=1}^{B} (\Delta_2^b - \Delta_2)(\Delta_2^b - \Delta_2)^\tau,
$$

respectively, which yields

$$
\widehat{h}_0 = -\left\{\frac{1}{B}\sum_{b=1}^{B} (\Delta_2^b - \Delta_2)(\Delta_2^b - \Delta_2)^\tau\right\}^{-1} \frac{1}{B}\sum_{b=1}^{B} (\widetilde{\gamma}_1^b - \widetilde{\gamma}_1)(\Delta_2^b - \Delta_2).
$$

Hence, our improved estimator for γ_1 becomes

$$
\widehat{\gamma}_1 = \widetilde{\gamma}_1 + \widehat{h}_0^\tau \Delta_2. \tag{6}
$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

When $f_k(\boldsymbol{X}_{t,k};\boldsymbol{\theta}_k) = \boldsymbol{\theta}_k^\tau \boldsymbol{X}_{t,k}$ for $k=1$ and 2, we have the following explicit formulas

$$\widetilde{\boldsymbol{\theta}}_k = (\frac{1}{T_k}\sum_{t=t_k}^{T_k} \boldsymbol{X}_{t,k}\boldsymbol{X}_{t,k}^\tau)^{-1}\frac{1}{T_k}\sum_{t=t_k}^{T_k} Y_{t,k}\boldsymbol{X}_{t,k} \text{ and } \widetilde{\boldsymbol{\theta}}_k^b = (\frac{1}{T_k}\sum_{t=t_k}^{T_k} \delta_t^b\boldsymbol{X}_{t,k}\boldsymbol{X}_{t,k}^\tau)^{-1}\frac{1}{T_k}\sum_{t=t_k}^{T_k} \delta_t^b Y_{t,k}\boldsymbol{X}_{t,k}$$

for $k=1$ and 2,

$$\Delta_2 = \frac{1}{T_1 - t_1\vee t_2 + 1}\sum_{t=t_1\vee t_2}^{T_1}(Y_{t,2} - \boldsymbol{X}_{t,2}^\tau\widetilde{\boldsymbol{\theta}}_2)\boldsymbol{X}_{t,2}, \text{ and } \Delta_2^b = \frac{\sum_{t=t_1\vee t_2}^{T_1}\delta_t^b(Y_{t,2} - \boldsymbol{X}_{t,2}^\tau\widetilde{\boldsymbol{\theta}}_2^b)\boldsymbol{X}_{t,2}}{\sum_{t=t_1\vee t_2}^{T_1}\delta_t^b}.$$

To establish the asymptotic limit of $\widehat{\gamma}_1$ and quantify its uncertainty, we use the following regularity conditions, where $\boldsymbol{\theta}_k^0$ is the true value of $\boldsymbol{\theta}_k, k=1,2$.

- C1) For $k=1$ and 2, we assume

$$\varepsilon_{t,k} = \sigma_{t,k}\eta_{t,k},\ \sigma_{t,k}^2 = w_k + \sum_{i=1}^{p_k} a_{i,k}\varepsilon_{t-i,k}^2 + \sum_{j=1}^{q_k} b_{j,k}\sigma_{t-j,k}^2$$

for some $w_k > 0, a_{1,k}\ge 0,\dots,a_{p_k,k}\ge 0, b_{1,k}\ge 0,\dots,b_{q_k,k}\ge 0$, where $\{(\eta_{t,1},\eta_{t,2})^\tau\}_{t=1}^{T_2}$ is a sample of independent and identically distributed random vectors with $E(\eta_{t,k})=0$ and $E(\eta_{t,k}^2)=1$. Further, assume

$$\sum_{i=1}^{p_k} a_{i,k} + \sum_{j=1}^{q_k} b_{j,k} < 1, \text{ for } k=1,2.$$

- C2) $\{\boldsymbol{X}_{t,k}\}_{t=1}^{T_2}$ is stationary and ergodic, and $E(\eta_{t,k}\,|\,\{\boldsymbol{X}_{s,1},\boldsymbol{X}_{s,2}\}_{s=1}^t) = \boldsymbol{0}$, for $k=1,2$.

- C3) For $k=1,2$ and $\boldsymbol{\theta}_k = (\theta_{k,1},\dots,\theta_{k,d_k})^\tau$, there exist a neighborhood Ω_k of $\boldsymbol{\theta}_k^0$ and $\delta_0 > 0$ such that for $1\le l_1,l_2,l_3\le d_k$,

$$E\sup_{\boldsymbol{\theta}_k\in\Omega_k}\left|\frac{\partial f_k(\boldsymbol{X}_{t,k};\boldsymbol{\theta}_k)}{\partial\theta_{k,l_1}}\frac{\partial^2 f_k(\boldsymbol{X}_{t,k};\boldsymbol{\theta}_k)}{\partial\theta_{k,l_2}\partial\theta_{k,l_3}}\right| < \infty,\ E\sup_{\boldsymbol{\theta}_k\in\Omega_k}\left|\varepsilon_{t,k}\frac{\partial^3 f_k(\boldsymbol{X}_{t,k};\boldsymbol{\theta}_k)}{\partial\theta_{k,l_1}\partial\theta_{k,l_2}\partial\theta_{k,l_3}}\right| < \infty,$$
$$E\sup_{\boldsymbol{\theta}_k\in\Omega_k}\left|\frac{\partial f_k(\boldsymbol{X}_{t,k};\boldsymbol{\theta}_k)}{\partial\theta_{k,l_1}}\frac{\partial f_k(\boldsymbol{X}_{t,k};\boldsymbol{\theta}_k)}{\partial\theta_{k,l_2}}\right|^{1+\delta_0} < \infty,\ E\sup_{\boldsymbol{\theta}_k\in\Omega_k}\left|\varepsilon_{t,k}\frac{\partial^2 f_k(\boldsymbol{X}_{t,k};\boldsymbol{\theta}_k)}{\partial\theta_{k,l_1}\partial\theta_{k,l_2}}\right|^{1+\delta_0} < \infty,$$
$$E\sup_{\boldsymbol{\theta}_k\in\Omega_k}\left|\varepsilon_{t,k}\frac{\partial f_k(\boldsymbol{X}_{t,k};\boldsymbol{\theta}_k)}{\partial\theta_{k,l_1}}\right|^{2+\delta_0} < \infty.$$

- C4) $\frac{T_1-(t_1\vee t_2)}{T_1-t_1}\to q\in(0,1]$ as $T_1-t_1\to\infty$.

We use the popular GARCH models introduced in Engle (1982) and Bollerslev (1986) in C1) for heteroscedasticity and allow general, cross-sectional dependence between $\eta_{t,1}$ and $\eta_{t,2}$. However, the proposed inference does not estimate the GARCH models and the cross-sectional dependence. Hence, it is easy to relax the GARCH errors to some general positively measurable functions $\sigma_{t,k} = g_k(\eta_{t-1,k}, \eta_{t-2,k}, \cdots)$ such as in the generalized GARCH models in Zhu and Ling (2015). Condition C2) holds when $\{\boldsymbol{X}_t\}$ is independent of $\{(\eta_{t,1}, \eta_{t,2})^\tau\}$. Condition C3) ensures that the least squares estimator is asymptotically normal. When f_k is linear in $\boldsymbol{X}_{t,k}$ for $k = 1$ and 2, the partial derivatives of f_k with order larger than one are zero, hence condition C3) becomes $E(|\varepsilon_{t,k}(1 + ||\boldsymbol{X}_{t,k}||)|^{2+\delta_0}) < \infty$ and $E(||\boldsymbol{X}_{t,k}||^{2+\delta_0}) < \infty$ for some $\delta_0 > 0$. Recall that we focus on the particular function G defined in (5) and $\widehat{\gamma}_1$ as given in (6).

Theorem 1. *Suppose model (1), with $T = T_2$, satisfies conditions C1)–C4). Then, as $T_1 - t_1 \to \infty$ and $B \to \infty$,*

$$\frac{\widetilde{\gamma}_1 - \gamma_1}{\widetilde{\sigma}_1} \xrightarrow{d} N(0,1) \text{ and } \frac{\widehat{\gamma}_1 - \gamma_1}{\widehat{\sigma}_1} \xrightarrow{d} N(0,1),$$

where

$$\widetilde{\sigma}_1^2 = \frac{1}{B}\sum_{b=1}^{B}(\widetilde{\gamma}_1^b - \widetilde{\gamma}_1)^2, \ \widehat{\sigma}_1^2 = \widetilde{\sigma}_1^2 - \widehat{h}_0^\tau \left\{ \frac{1}{B}\sum_{b=1}^{B}(\Delta_2^b - \Delta_2)(\Delta_2^b - \Delta_2)^\tau \right\} \widehat{h}_0.$$

In the simulation study section and the application section, it will become clear that the variance reduction when using $\widehat{\gamma}_1$ instead of $\widetilde{\gamma}_1$ can be sizable.

Remark 1. *In case a third time series $\{Y_{t,3}, \boldsymbol{X}_{t,3}\}_{t=t_3}^{T_3}$ is observed, which overlaps or has a different length from the first sequence, we can consider the class of estimators*

$$\widetilde{\gamma}_1 + h^\tau \Delta_2 + h_1^\tau \Delta_3,$$

where Δ_3 is constructed similarly as Δ_2 for some known vector of functions G_1 (instead of G). As before, we choose h and h_1 by minimizing the asymptotic variance of the above estimators. Then an extension of Theorem 1 could be derived under appropriate assumptions. Adding the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

third time series improves the performance of the new estimator, in particular, if the first and second time series as well as the first and third time series are strongly dependent, whereas the second and third time series are less dependent. The case of three or more time series is beyond the scope of this paper and will not be pursued. See Ahmed and Einmahl (2019), Section 3, for results along these lines in extreme value statistics.

Remark 2. *When $\gamma_1 = K(\boldsymbol{\theta}_1)$ is a vector with K a known function from $\mathbb{R}^{d_1}$ to $\mathbb{R}^{d_3}$, we can still consider the class of estimators $\widetilde{\gamma}_1 + h^\tau \Delta_2$ and find an optimal h by minimizing*

$$E\{(\widetilde{\gamma}_1 - \gamma_1 + h^\tau \Delta_2)^\tau (\widetilde{\gamma}_1 - \gamma_1 + h^\tau \Delta_2)\} = E\{(\widetilde{\gamma}_1 - \gamma_1 + h^\tau \Delta_2 \mathbf{1}_{d_3})^\tau (\widetilde{\gamma}_1 - \gamma_1 + h^\tau \Delta_2 \mathbf{1}_{d_3})\},$$

where $\mathbf{1}_{d_3}$ denotes a d_3-vector with all elements being one, which leads to

$$h_0 = -\{E(\Delta_2 \Delta_2^\tau)\}^{-1} E\{(\mathbf{1}_{d_3}^\tau (\widetilde{\gamma}_1 - \gamma_1))\Delta_2\}$$

and the new estimator $\widetilde{\gamma}_1 + \hat{h}_0^\tau \Delta_2$, where $\hat{h}_0$ is an estimator of h_0 via the same random weighted bootstrap method as in Theorem 1. More effectively but tediously, we can minimize the asymptotic variance of the class of estimators $\widetilde{\gamma}_1 + H\Delta_2$ with H being a matrix instead of a vector.

2.2 Correlated errors

Next, we generalize the method to correlated and heteroscedastic errors by considering the following models with autoregressive (AR) errors:

$$Y_{t,k} = f_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k) + U_{t,k},\ U_{t,k} = \sum_{j=1}^{s_k} \phi_{j,k} U_{t-j,k} + \varepsilon_{t,k} \tag{7}$$

for $k = 1, 2$, where the $\varepsilon_{t,k}$ satisfy C1) as before. Put $\boldsymbol{\phi}_k = (\phi_{1,k}, \ldots, \phi_{s_k,k})^\tau$. Like in Hall and Yao (2003) and Liu, Chen and Yao (2010), we can take the AR structure into account to estimate $\boldsymbol{\theta}_k$ by

$$(\boldsymbol{\theta}_k^*, \boldsymbol{\phi}_k^*) = \arg\min_{\boldsymbol{\theta}_k, \boldsymbol{\phi}_k} \sum_{t=t_k}^{T_k} \{Y_{t,k} - f_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k) - \sum_{j=1}^{s_k} \phi_{j,k}(Y_{t-j,k} - f_k(\boldsymbol{X}_{t-j}; \boldsymbol{\theta}_k))\}^2. \tag{8}$$

To improve the estimator $\gamma_1^* = K(\boldsymbol{\theta}_1^*)$ with the aid of the second series, we minimize the mean squared error within the class of estimators

$$\gamma_1^* + h^\tau \Delta_2^*,$$

where

$$\Delta_2^* = \frac{1}{T_1 - (t_1 \vee t_2) + 1} \sum_{t=t_1 \vee t_2}^{T_1} \frac{\partial}{\partial \boldsymbol{\theta}_2} \left\{ Y_{t,2} - f_2(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2^*) - \sum_{j=1}^{s_2} \phi_{j,2}^* (Y_{t-j,2} - f_2(\boldsymbol{X}_{t-j}; \boldsymbol{\theta}_2^*)) \right\}^2.$$

As before, we can use the following random weight bootstrap method to estimate the optimal h and obtain the new estimator:

- Step Bi) Draw a random sample with size T_2 from a probability distribution with mean one and variance one, say the standard exponential distribution. Denote it by $\{\delta_t^b\}_{t=1}^{T_2}$.

- Step Bii) Compute

$$(\boldsymbol{\theta}_1^{*b}, \boldsymbol{\phi}_1^{*b}) = \arg\min_{\boldsymbol{\theta}_1, \boldsymbol{\phi}_1} \sum_{t=t_1}^{T_1} \delta_t^b \{Y_{t,1} - f_1(\boldsymbol{X}_{t,1}; \boldsymbol{\theta}_1) - \sum_{j=1}^{s_1} \phi_{j,1}(Y_{t-j,1} - f_1(\boldsymbol{X}_{t-j,1}; \boldsymbol{\theta}_1))\}^2,$$
$$\gamma_1^{*b} = K(\boldsymbol{\theta}_1^{*b}),$$
$$(\boldsymbol{\theta}_2^{*b}, \boldsymbol{\phi}_2^{*b}) = \arg\min_{\boldsymbol{\theta}_2, \boldsymbol{\phi}_2} \sum_{t=t_2}^{T_2} \delta_t^b \{Y_{t,2} - f_2(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2) - \sum_{j=1}^{s_2} \phi_{j,2}(Y_{t-j,2} - f_2(\boldsymbol{X}_{t-j,2}; \boldsymbol{\theta}_2))\}^2,$$
$$\Delta_2^{*b} = \frac{\sum_{t=t_1 \vee t_2}^{T_1} \delta_t^b \frac{\partial}{\partial \boldsymbol{\theta}_2} \left\{ Y_{t,2} - f_2(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2^{*b}) - \sum_{j=1}^{s_2} \phi_{j,2}^{*b} (Y_{t-j,2} - f_2(\boldsymbol{X}_{t-j}; \boldsymbol{\theta}_2^{*b})) \right\}^2}{\sum_{t=t_1 \vee t_2}^{T_1} \delta_t^b}.$$

- Step Biii) Repeat the above two steps B times to get $\{\gamma_1^{*b}, \Delta_2^{*b}\}_{b=1}^B$.

Therefore,

$$h_0^* = -\left\{ \frac{1}{B} \sum_{b=1}^B (\Delta_2^{*b} - \Delta_2^*)(\Delta_2^{*b} - \Delta_2^*)^\tau \right\}^{-1} \frac{1}{B} \sum_{b=1}^B (\gamma_1^{*b} - \gamma_1^*)(\Delta_2^{*b} - \Delta_2^*),$$

and the proposed estimator for γ_1 becomes

$$\widehat{\gamma}_1^* = \gamma_1^* + h_0^{*\tau} \Delta_2^*.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

Theorem 2. *Suppose model (7) satisfies conditions C1)–C4) with the* $\{U_{t,k}\}$ *being strictly stationary. Then, as* $T_1 - t_1 \to \infty$ *and* $B \to \infty$,

$$\frac{\gamma_1^* - \gamma_1}{\sigma_1^*} \xrightarrow{d} N(0,1) \text{ and } \frac{\widehat{\gamma}_1^* - \gamma_1}{\widehat{\sigma}_1^*} \xrightarrow{d} N(0,1),$$

where

$$\sigma_1^{*2} = \frac{1}{B}\sum_{b=1}^{B}(\gamma_1^{*b} - \gamma_1^*)^2 \text{ and } \widehat{\sigma}_1^{*2} = \sigma_1^{*2} - h_0^{*\tau}\left\{\frac{1}{B}\sum_{b=1}^{B}(\Delta_2^{*b} - \Delta_2^*)(\Delta_2^{*b} - \Delta_2^*)^\tau\right\} h_0^*.$$

An application is the performance evaluation of mutual funds, where researchers and practitioners often use the one-factor model in Jensen (1968), the three-factor model in Fama and French (1996), or the four-factor model in Carhart (1997) to model the fund returns:

$$\begin{cases} Y_{t,k} = \alpha_k + \boldsymbol{\beta}_k^\tau \boldsymbol{X}_t + U_{t,k},\ U_{t,k} = \sum_{j=1}^{s_k} \phi_{j,k} U_{t-j,k} + \varepsilon_{t,k}, \\ \varepsilon_{t,k} = \sigma_{t,k}\eta_{t,k},\ \sigma_{t,k}^2 = w_k + \sum_{i=1}^{p_k} a_{i,k}\varepsilon_{t-i,k}^2 + \sum_{j=1}^{q_k} b_{j,k}\sigma_{t-j,k}^2, \end{cases} \tag{9}$$

where $\alpha_k \in \mathbb{R}$, $\boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,d})^\tau \in \mathbb{R}^d$, $\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,d})^\tau$ and $X_{t,1}$ denotes the market excess return. In this case, α_k measures the stock picking skill for the kth fund, and $\beta_{k,1}$ $(> 1, = 1, < 1)$ represents the fund manager's risk attitude (risk loving, neutral, averse). Often least squares inference is employed for α_k and $\boldsymbol{\beta}_k$. In practice, funds are usually observed in different windows and cross-sectionally dependent, and some funds have a short time series. Hence, it is important to improve the statistical inference for such funds with a short series using funds with longer time series. In the data analysis below, we consider models (1) and (7) with $f_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k) = \alpha_k + \boldsymbol{\beta}_k^\tau \boldsymbol{X}_t$, $\boldsymbol{\theta}_k = (\alpha_k, \boldsymbol{\beta}_k^\tau)^\tau$, and apply the improved inference to some funds with a short time window (i.e., $k = 1$ in the model) in a US mutual funds dataset with daily returns from September 1, 1998 to December 31, 2018. The longer series (i.e., $k = 2$ in the model) consists of the average daily returns of 654 funds.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

3 Simulation Study

This section examines the finite sample performance of the improved estimation method. To better appreciate the applicability, we study the following settings adopted from the dataset analyzed in the next section.

We draw 10000 random samples from the one-factor model with uncorrelated, heteroscedastic errors

$$\begin{cases} Y_{t,k} = \alpha_k + \beta_k X_t + \varepsilon_{t,1},\ \varepsilon_{t,k} = \sigma_{t,k}\eta_{t,k},\ \sigma_{t,k}^2 = w_k + a_k\varepsilon_{t-1,k}^2 + b_k\sigma_{t-1,k}^2, \\ X_t = \bar{\mu} + \bar{\phi}X_{t-1} + \bar{\varepsilon}_t,\ \bar{\varepsilon}_t = \bar{\sigma}_t\bar{\eta}_t,\ \bar{\sigma}_t^2 = \bar{w} + \bar{a}\bar{\varepsilon}_{t-1}^2 + \bar{b}\bar{\sigma}_{t-1}^2. \end{cases}$$

We use the fund with the fund's identifier wficn=601187 and the largest sample size 5116 in our data set to get

$$\alpha_2 = 0.0082,\ \beta_2 = 0.9190,\ w_2 = 0.0027,\ a_2 = 0.0606,\ b_2 = 0.9312,$$

$$\bar{\mu} = 0.0772,\ \bar{\phi} = -0.0386,\ \bar{w} = 0.0102,\ \bar{a} = 0.0974,\ \bar{b} = 0.9000,$$

and the fund with the fund's identifier wficn=105877 and sample size 254 to get

$$\alpha_1 = -0.0194,\ \beta_1 = 0.8063,\ w_1 = 0.0103,\ a_1 = 0.0915,\ b_1 = 0.8768.$$

We draw the $(\eta_{t,1}, \eta_{t,2})$ from a bivariate normal distribution with means zero, variances one, and correlation coefficient $\rho = 0.3, 0.4, 0.5$ and independently draw the $\bar{\eta}_t$ from the $N(0,1)$-distribution. We consider if the underlying first series is completely inside or only overlaps with the other longer series:

- Case i) We observe the second series from $t = 1, \ldots, 5000$ and the first series from $t = 1, \ldots, 5000p$ with $p = 0.05, 0.1, 0.3, 0.5$. Hence, the first series is inside the second one.

- Case ii) We observe the first series from $t = 1, \ldots, 1000$ and the second one from $t = 1000p + 1, \ldots, 1000p + 5000$ with $p = 0.25, 0.5, 0.75$. Hence, the first series overlaps with the second one.

For implementing the new inference method, we use $B = 10000$ in the random weight bootstrap method.

Using the aforementioned 10000 random samples, we compute the simulated means of $\widetilde{\alpha}_1$ and $\widehat{\alpha}_1$ to examine the performance, the ratio of the simulated variance of $\widehat{\alpha}_1$ to that of $\widetilde{\alpha}_1$ to see the efficiency improvement, the average of the ratios of the estimated variance of $\widehat{\alpha}_1$ to its simulated variance to check the accuracy of the proposed random weight bootstrap method for uncertainty quantification. We also calculate similar quantities for $\widetilde{\beta}_1$ and $\widehat{\beta}_1$ based on these 10000 repetitions. Tables 1 and 2 report these quantities for Cases i) and ii), respectively.

The results for the averages of the estimators show that the proposed estimators and the least squares estimators have similar small biases, but, more importantly, the results in the columns of $V_{\widehat{\alpha}_1}/V_{\widetilde{\alpha}_1}$ and $V_{\widehat{\beta}_1}/V_{\widetilde{\beta}_1}$ show that the new estimators have substantially smaller variances than the least squares estimators, especially when the two series are moderately correlated ($\rho = 0.5$) the variance reduction can increase to 23%. The averages of the ratios of the estimated variances to the simulated variances support that the random weight bootstrap method performs well for quantifying the estimation uncertainty.

4 Analysis of Mutual Funds

We consider the daily returns of US mutual funds from September 1, 1998 to December 31, 2018. We look at funds with the largest sample size of 5116 and smaller sample sizes of 254, 506, 1025, 1534, and 2053, which are about 0.05, 0.1, 0.2, 0.3, and 0.4, of the largest sample size. The fund identifier is “wficn”. The number of funds with the largest sample size is 654. We use the average daily returns of these 654 funds as our $Y_{t,2}$ and the daily returns of each of the funds with a smaller sample size as our $Y_{t,1}$. We employ the one-factor model in Jensen (1968) with GARCH(1,1) errors and both $X_{t,1} = X_{t,2}$ being the market excess return. That is,

Table 1: Case i)

p	$\widetilde{\alpha}_1$	$\widehat{\alpha}_1$	$V_{\widehat{\alpha}_1}/V_{\widetilde{\alpha}_1}$	$\widehat{V}_{\widehat{\alpha}_1}/V_{\widehat{\alpha}_1}$	$\widetilde{\beta}_1$	$\widehat{\beta}_1$	$V_{\widehat{\beta}_1}/V_{\widetilde{\beta}_1}$	$\widehat{V}_{\widehat{\beta}_1}/V_{\widehat{\beta}_1}$
				$\rho = 0.3$				
0.05	−0.0189	−0.0188	0.9187	0.9764	0.8067	0.8067	0.9439	0.9414
0.1	−0.0192	−0.0193	0.9275	0.9915	0.8063	0.8063	0.9382	0.9666
0.3	−0.0191	−0.0191	0.9553	1.0051	0.8063	0.8063	0.9357	0.9778
0.5	−0.0192	0.0193	0.9701	1.0200	0.8061	0.8061	0.9568	0.9811
				$\rho = 0.4$				
0.05	−0.0189	−0.0188	0.8529	0.9822	0.8066	0.8066	0.8772	0.9457
0.1	−0.0193	−0.0193	0.8690	0.9920	0.8063	0.8063	0.8794	0.9694
0.3	−0.0191	−0.0191	0.9126	1.0070	0.8063	0.8063	0.8857	0.9806
0.5	−0.0192	−0.0193	0.9401	1.0245	0.8061	0.8061	0.9215	0.9813
				$\rho = 0.5$				
0.05	−0.0190	−0.0188	0.7704	0.9837	0.8066	0.8066	0.7940	0.9494
0.1	−0.0192	−0.0192	0.7963	0.9866	0.8061	0.8062	0.7991	0.9798
0.3	−0.0191	−0.0191	0.8553	1.0086	0.8063	0.8063	0.8212	0.9839
0.5	−0.0192	−0.0192	0.8961	1.0414	0.8061	0.8061	0.8736	0.9924

Based on 10000 repetitions, this table reports the averages of the estimators $(\widetilde{\alpha}_1, \widehat{\alpha}_1, \widetilde{\beta}_1, \widehat{\beta}_1)$, the ratios of the simulated variances of the improved estimators to the least squares estimators $(V_{\widehat{\alpha}_1}/V_{\widetilde{\alpha}_1}, V_{\widehat{\beta}_1}/V_{\widetilde{\beta}_1})$, and the average ratios of the estimated variances to the simulated variances of $\widehat{\alpha}_1$ and $\widehat{\beta}_1$, that is, $\widehat{V}_{\widehat{\alpha}_1}/V_{\widehat{\alpha}_1}$ and $\widehat{V}_{\widehat{\beta}_1}/V_{\widehat{\beta}_1}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

Table 2: Case ii)

p	$\widetilde{\alpha}_1$	$\widehat{\alpha}_1$	$V_{\widehat{\alpha}_1}/V_{\widetilde{\alpha}_1}$	$\widehat{V}_{\widehat{\alpha}_1}/V_{\widehat{\alpha}_1}$	$\widetilde{\beta}_1$	$\widehat{\beta}_1$	$V_{\widehat{\beta}_1}/V_{\widetilde{\beta}_1}$	$\widehat{V}_{\widehat{\beta}_1}/V_{\widehat{\beta}_1}$
				$\rho = 0.3$				
0.25	−0.0195	−0.0195	0.9472	1.0257	0.8062	0.8062	0.9526	0.9522
0.5	−0.0195	−0.0195	0.9648	0.9935	0.8060	0.8061	0.9673	0.9933
0.75	−0.0190	−0.0190	0.9796	1.0056	0.8062	0.8062	0.9881	0.9665
				$\rho = 0.4$				
0.25	−0.0195	−0.0195	0.9041	1.0263	0.8062	0.8062	0.9093	0.9510
0.5	−0.0195	−0.0195	0.9349	0.9930	0.8060	0.8061	0.9365	0.9907
0.75	−0.0190	−0.0190	0.9628	1.0072	0.8062	0.8062	0.9723	0.9660
				$\rho = 0.5$				
0.25	−0.0195	−0.0195	0.8506	1.0239	0.8062	0.8062	0.8565	0.9441
0.5	−0.0195	−0.0195	0.8959	0.9930	0.8060	0.8061	0.8966	0.9874
0.75	−0.0190	−0.0190	0.9414	1.0086	0.8062	0.8061	0.9515	0.9655

Based on 10000 repetitions, this table reports the averages of the estimators $(\widetilde{\alpha}_1, \widehat{\alpha}_1, \widetilde{\beta}_1, \widehat{\beta}_1)$, the ratios of the simulated variances of the improved estimators to the least squares estimators $(V_{\widehat{\alpha}_1}/V_{\widetilde{\alpha}_1}, V_{\widehat{\beta}_1}/V_{\widetilde{\beta}_1})$, and the average ratios of the estimated variances to the simulated variances of $\widehat{\alpha}_1$ and $\widehat{\beta}_1$, that is, $\widehat{V}_{\widehat{\alpha}_1}/V_{\widehat{\alpha}_1}$ and $\widehat{V}_{\widehat{\beta}_1}/V_{\widehat{\beta}_1}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

we consider

$$Y_{t,k} = \alpha_k + \beta_k X_t + \varepsilon_{t,k},\ \varepsilon_{t,k} = \sigma_{t,k}\eta_{t,k},\ \sigma^2_{t,k} = w_k + a_k\varepsilon^2_{t-1,k} + b_k\sigma^2_{t-1,k}.$$

We use $B = 10000$ in the random weight bootstrap method and the 'fGarch' R package to fit the GARCH(1,1) models to the fitted residuals of the one-factor model and estimate $E(\eta_{t,1}\eta_{t,2})$ by the sample mean. We only report funds where this average exceeds 0.2 in absolute value. Note that we do not estimate the GARCH errors when we study the least squares estimation and improved inference for α_k and β_k.

Table 3 presents the sample mean corresponding to $E(\eta_{t,1}\eta_{t,2})$, the least squares estimators and the improved ones, and the ratios of estimated variances. We see a variance reduction for all cases; the reduction can be as large as 24%. Table 4 reports 90% confidence intervals based on the improved estimators and the least squares estimators using the random weight bootstrap method for estimating the asymptotic variances. Using the confidence intervals, we summarize the different conclusions from these two methods:

- For fund wficn=401297 with size 254, least squares estimation suggests $\alpha_1 = 0$ and $\beta_1 = 0$, whereas the improved inference shows $\alpha_1 > 0$ and $\beta_1 > 0$.
- For fund wficn=101466 with size 506, least squares estimation shows risk averse ($\beta_1 < 1$), whereas the improved inference suggests risk neutral ($\beta_1 = 1$).
- For funds wficn=105279, and wficn=106044 with size 506, the least squares estimation suggests $\alpha_1 = 0$, but the improved inference shows $\alpha_1 < 0$.
- For fund wficn=410526 with size 1025, least squares estimation shows $\alpha_1 < 0$, but the improved inference suggests $\alpha_1 = 0$.
- For fund wficn=400286 with size 2053, the least squares estimation shows risk loving ($\beta_1 > 1$), but the improved inference suggests risk neutral.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

5 Conclusion

Financial returns/losses from different business lines or institutions may be observed in different but overlapping time windows with cross-sectional dependence. Hence, after modeling each series of losses by a parametric regression, the two regression errors are correlated and overlapping. In this setting, we present an improved inference method over the standard least squares estimation by combining the least squares estimator from the underlying series and the score functions of another series computed for the overlapping period. A simulation study shows that the improvement is already substantial when the two regression errors are moderately correlated. Applying mutual funds reveals that the novel estimation method frequently leads to different conclusions on fund managers' skills and risk attitudes.

Appendix: Proofs

Proof Theorem 1. Define $m_1 = T_1 - t_1 + 1$, $m_2 = T_2 - t_2 + 1$, $m_{12} = T_1 - (t_1 \vee t_2) + 1$,

$$\Gamma_{k\ell} = E\{\dot{f}_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k)\dot{f}_\ell^\tau(\boldsymbol{X}_{t,\ell}; \boldsymbol{\theta}_\ell)\}, \text{ and } \Sigma_{k\ell} = E\{\varepsilon_{t,k}\dot{f}_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k)\varepsilon_{t,\ell}\dot{f}_\ell^\tau(\boldsymbol{X}_{t,\ell}; \boldsymbol{\theta}_\ell)\}.$$

Then, it can be shown using a Taylor expansion, conditions C1)–C4), and the weak law of large numbers and central limit theorem for martingale differences in Hall and Heyde (1980) that for $k = 1, 2$

$$\sqrt{m_k}(\widetilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k) = \Gamma_{kk}^{-1} \frac{1}{\sqrt{m_k}} \sum_{t=t_k}^{T_k} \varepsilon_{t,k}\dot{f}_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k) + o_p(1),$$

implying that

$$\begin{aligned} \sqrt{m_1}(\widetilde{\gamma}_1 - \gamma_1) &= \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1} \tfrac{1}{\sqrt{m_1}} \textstyle\sum_{t=t_1}^{T_1} \varepsilon_{t,1}\dot{f}_1(\boldsymbol{X}_{t,1}; \boldsymbol{\theta}_1) + o_p(1) \\ &\xrightarrow{d} N\left(0, \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1}\Sigma_{11}\Gamma_{11}^{-1}\dot{K}(\boldsymbol{\theta}_1)\right). \end{aligned} \tag{10}$$

Noting that

$$E\{\varepsilon_{s,1}\varepsilon_{t,2}\dot{f}_1(\boldsymbol{X}_{s,1}; \boldsymbol{\theta}_1)\dot{f}_2^\tau(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2)\} = \boldsymbol{0} \text{ for } s \neq t,$$

Table 3: US mutual funds analysis: estimates

wficn	size	$E(\eta_{t,1}\eta_{t,2})$	$\widetilde{\alpha}_1$	$\widehat{\alpha}_1$	$\widehat{V}_{\widehat{\alpha}_1}/\widehat{V}_{\widetilde{\alpha}_1}$	$\widetilde{\beta}_1$	$\widehat{\beta}_1$	$\widehat{V}_{\widehat{\beta}_1}/\widehat{V}_{\widetilde{\beta}_1}$
105877	254	−0.3335	−0.0194	−0.0324	0.8612	0.8063	0.7571	0.8614
401297	254	0.3388	0.0634	0.1340	0.8758	0.0896	0.2080	0.9105
101466	506	0.4625	0.0758	0.1170	0.8124	0.7862	0.9424	0.8772
101588	506	0.3533	0.0480	0.1020	0.9324	1.1247	1.2130	0.9648
101589	506	0.4141	0.0520	0.0224	0.8841	1.2393	1.3408	0.9322
105279	506	−0.3818	−0.0257	−0.0441	0.8814	0.9928	0.9772	0.9719
105704	506	0.5228	0.0351	0.0121	0.8262	0.9249	0.9882	0.9870
106044	506	0.4636	−0.0158	−0.0335	0.8056	0.5618	0.6012	0.9484
401173	506	0.4537	0.0074	0.0456	0.8172	0.3570	0.4781	0.8755
105511	1025	0.2975	0.0242	0.0025	0.9345	0.2531	0.3200	0.9651
107563	1025	−0.4882	−0.0062	−0.0070	0.9030	0.9631	0.9540	0.9857
410526	1025	0.3754	−0.0092	−0.0041	0.8953	0.7924	0.7786	0.7614
500982	1534	−0.3515	−0.0071	−0.0094	0.9194	1.0585	1.0656	0.9888
226946	2053	0.4260	−0.0067	−0.0076	0.8481	0.8919	0.8887	0.9152
400286	2053	0.5349	−0.0003	−0.0019	0.8098	1.0361	1.0010	0.8842

Using the one-factor model and $B = 10000$ in the random weight bootstrap method, this table reports estimators of $\widetilde{\alpha}_1, \widehat{\alpha}_1, \widetilde{\beta}_1, \widehat{\beta}_1$, the ratios of estimated variances ($\widehat{V}_{\widehat{\alpha}_1}/\widehat{V}_{\widetilde{\alpha}_1}$ and $\widehat{V}_{\widehat{\beta}_1}/\widehat{V}_{\widetilde{\beta}_1}$), and the sample mean corresponding to $E(\eta_{t,1}\eta_{t,2})$ after fitting GARCH(1,1) models to the first and second residuals.

Table 4: US mutual funds analysis: 90% confidence intervals

wficn	size	$\widetilde{\alpha}_1$	$\widehat{\alpha}_1$	$\widetilde{\beta}_1$	$\widehat{\beta}_1$
105877	254	(−0.0748, 0.0360)	(−0.0838, 0.0190)	(0.7629, 0.8496)	(0.7168, 0.7973)
401297	254	*(−0.0782, 0.2051)*	*(0.0015, 0.2665)*	*(−0.0264, 0.2056)*	*(0.0973, 0.3186)*
101466	506	(0.0074, 0.1441)	(0.0553, 0.1786)	*(0.7185, 0.8539)*	*(0.8790, 1.0058)*
101588	506	(−0.0659, 0.1620)	(−0.0081, 0.2120)	(1.0419, 1.2076)	(1.1317, 1.2944)
101589	506	(−0.0145, 0.1185)	(−0.0401, 0.0850)	(1.1868, 1.2918)	(1.2901, 1.3915)
105279	506	*(−0.0611, 0.0097)*	*(−0.0773, −0.0109)*	(0.9627, 1.0228)	(0.9426, 1.0019)
105704	506	(−0.0473, 0.1176)	(−0.0628, 0.0871)	(0.8463, 1.0036)	(0.9101, 1.0664)
106044	506	*(−0.0441, 0.0125)*	*(−0.0589, −0.0081)*	(0.5386, 0.5851)	(0.5786, 0.6238)
401173	506	(−0.0586, 0.0716)	(−0.0125, 0.1036)	(0.3015, 0.4125)	(0.4261, 0.5300)
105511	1025	(−0.0136, 0.0621)	(−0.0341, 0.0391)	(0.2196, 0.2867)	(0.2871, 0.3530)
107563	1025	(−0.0187, 0.0062)	(−0.0188, 0.0049)	(0.9516, 0.9747)	(0.9426, 0.9655)
410526	1025	*(−0.0177, −0.0007)*	*(−0.0122, 0.0039)*	(0.7749, 0.8101)	(0.7631, 0.7940)
500982	1534	(−0.0175, 0.0033)	(−0.0193, 0.0006)	(1.0468, 1.0702)	(1.0539, 1.0772)
226946	2053	(−0.0190, 0.0057)	(−0.0190, 0.0037)	(0.8765, 0.9072)	(0.8740, 0.9034)
400286	2053	(−0.0132, 0.0127)	(−0.0136, 0.0097)	*(1.0191, 1.0531)*	*(0.9850, 1.0170)*

This table reports 90% confidence intervals based on $\widetilde{\alpha}_1, \widehat{\alpha}_1, \widetilde{\beta}_1, \widehat{\beta}_1$ using the one-factor model and $B = 10000$ in the random weight bootstrap method.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0371

and, as already indicated in (4),

$$\Delta_2 = \frac{1}{m_{12}} \sum_{t=t_1 \vee t_2}^{T_1} \varepsilon_{t,2} \dot{f}_2(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2) - \frac{1}{m_2} \sum_{t=t_2}^{T_2} \varepsilon_{t,2} \dot{f}_2(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2) + o_p(1/\sqrt{m_{12}}),$$

it can be shown that

$$\begin{aligned}
& E\{m_1(\widetilde{\gamma}_1 - \gamma_1)\Delta_2^\tau\} \\
&= \left\{\frac{1}{m_{12}} - \frac{1}{m_2}\right\} \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1} \sum_{t=t_1 \vee t_2}^{T_1} E\{\varepsilon_{t,1}\varepsilon_{t,2}\dot{f}_1(\boldsymbol{X}_{t,1}; \boldsymbol{\theta}_1)\dot{f}_2^\tau(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2)\} + o(1) \\
&= \left\{1 - \frac{m_{12}}{m_2}\right\} \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1} E\{\varepsilon_{t,1}\varepsilon_{t,2}\dot{f}_1(\boldsymbol{X}_{t,1}; \boldsymbol{\theta}_1)\dot{f}_2^\tau(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2)\} + o(1) \qquad (11) \\
&= \left\{1 - \frac{m_{12}}{m_2}\right\} \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1}\Sigma_{12} + o(1).
\end{aligned}$$

Similarly, for $k = 1, 2$

$$\sqrt{m_k}(\widetilde{\boldsymbol{\theta}}_k^b - \boldsymbol{\theta}_k) = \Gamma_{kk}^{-1} \frac{1}{\sqrt{m_k}} \sum_{t=t_k}^{T_k} \delta_t^b \varepsilon_{t,k} \dot{f}_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k) + o_p(1),$$

and hence,

$$\sqrt{m_k}(\widetilde{\boldsymbol{\theta}}_k^b - \widetilde{\boldsymbol{\theta}}_k) = \Gamma_{kk}^{-1} \frac{1}{\sqrt{m_k}} \sum_{t=t_k}^{T_k} (\delta_t^b - 1) \varepsilon_{t,k} \dot{f}_k(\boldsymbol{X}_{t,k}; \boldsymbol{\theta}_k) + o_p(1).$$

This yields that

$$\sqrt{m_1}(\widetilde{\gamma}_1^b - \widetilde{\gamma}_1) = \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1} \frac{1}{\sqrt{m_1}} \sum_{t=t_1}^{T_1} (\delta_t^b - 1)\varepsilon_{t,1}\dot{f}_1(\boldsymbol{X}_{t,1}; \boldsymbol{\theta}_1) + o_p(1), \qquad (12)$$

$$\Delta_2^b - \Delta_2 = \frac{1}{m_{12}} \sum_{t=t_1 \vee t_2}^{T_2} (\delta_t^b - 1)\varepsilon_{t,2}\dot{f}_2(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2) - \frac{1}{m_2} \sum_{t=t_2}^{T_2} (\delta_t^b - 1)\varepsilon_{t,2}\dot{f}_2(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2) + o_p(1/\sqrt{m_{12}}), \quad (13)$$

$$\begin{aligned}
& E\{m_1(\widetilde{\gamma}_1^b - \widetilde{\gamma}_1)(\Delta_2^b - \Delta_2)^\tau\} \\
&= \left\{\frac{1}{m_{12}} - \frac{1}{m_2}\right\} \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1} \sum_{t=t_1 \vee t_2}^{T_1} E\{(\delta_t^b - 1)^2 \varepsilon_{t,1}\varepsilon_{t,2}\dot{f}_1(\boldsymbol{X}_{t,1}; \boldsymbol{\theta}_1)\dot{f}_2^\tau(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2)\} + o(1) \\
&= \left\{1 - \frac{m_{12}}{m_2}\right\} \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1} E\{\varepsilon_{t,1}\varepsilon_{t,2}\dot{f}_1(\boldsymbol{X}_{t,1}; \boldsymbol{\theta}_1)\dot{f}_2^\tau(\boldsymbol{X}_{t,2}; \boldsymbol{\theta}_2)\} + o(1) \\
&= \left\{1 - \frac{m_{12}}{m_2}\right\} \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1}\Sigma_{12} + o(1). \qquad (14)
\end{aligned}$$

Using (10)–(14), we can show that

$$m_1\widetilde{\sigma}_1^2 = \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1}\Sigma_{11}\Gamma_{11}^{-1}\dot{K}(\boldsymbol{\theta}_1) + o_p(1),$$

$$\frac{1}{B}\sum_{b=1}^{B} m_1(\widetilde{\gamma}_1^b-\widetilde{\gamma}_1)(\Delta_2^b-\Delta_2)^\tau$$
$$=(1-\frac{m_{12}}{m_2})E\{\dot{K}^\tau(\widetilde{\boldsymbol{\theta}}_1)\Gamma_1^{-1}\varepsilon_{t,1}\varepsilon_{t,2}\dot{f}_1(\boldsymbol{X}_{t,1};\boldsymbol{\theta}_2)\dot{f}_2^\tau(\boldsymbol{X}_{t,2};\boldsymbol{\theta}_2)\}+o_p(1),$$
$$=m_1E\{(\widetilde{\gamma}_1-\gamma_1)\Delta_2\}+o_p(1),$$
$$\frac{m_1}{B}\sum_{b=1}^{B}(\Delta_2^b-\Delta_2)(\Delta_2^b-\Delta_2)^\tau=m_1E\{\Delta_2\Delta_2^\tau\}+o_p(1),\quad \widehat{h}_0=h_0+o_p(1),$$

and

$$\begin{aligned} m_1\widehat{\sigma}_1^2 &= m_1E(\widetilde{\gamma}_1-\gamma_1)^2-h_0^\tau m_1E\{\Delta_2\Delta_2^\tau\}h_0+o_p(1)\\ &= \dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1}\Sigma_{11}\Gamma_{11}^{-1}\dot{K}(\boldsymbol{\theta}_1) \\ &\quad -(1-\tfrac{m_{12}}{m_2})^2\dot{K}^\tau(\boldsymbol{\theta}_1)\Gamma_{11}^{-1}\Sigma_{12}\{E(m_1\Delta_2\Delta_2^\tau)\}^{-1}\Sigma_{12}^\tau\Gamma_{11}^{-1}\dot{K}(\boldsymbol{\theta}_1)+o_p(1). \end{aligned} \tag{15}$$

Hence, the theorem follows. □

The proof of Theorem 2 can be given along the same lines as that of Theorem 1 and will therefore be omitted.

Acknowledgments. We thank an associate editor and two reviewers for their helpful comments. John Einmahl held the Arie Kapteyn Chair 2019–2022 and gratefully acknowledges the corresponding research support. Peng's research was partly supported by the NSF grant of DMS-2012448.

References

Ahmed, H. and Einmahl, J.H.J. (2019). Improved estimation of the extreme value index using related variables. *Extremes 22, 553–569.*

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31, 307–327.*

Carhart, M.M. (1997). On persistence in mutual fund performance. *Journal of Finance 52, 57–82.*

Davison, A. C. (2011). *Statistical Models.* Cambridge University Press.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50, 987–1007.*

Fama, E.F. and French, K.R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51, 55–84.*

Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Application.* Academic Press.

Hall, P. and Yao, Q. (2003). Data tilting for time series. *Journal of the Royal Statistical Society Series B 65, 425–442.*

Jensen, M.C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance 23, 389–416.*

Jin, Z., Ying, Z. and Wei, L.J. (2001). A simple resampling method by perturbing the minimand. *Biometrika 88, 381-390.*

Liu, J.M., Chen, R. and Yao, Q. (2010). Nonparametric transfer function models. *Journal of Econometrics 157, 151–164.*

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics 21, 255–285.*

Patton, A.J. (2006). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics 21, 147–173.*

Zhang, A., Brown, L.D. and Cai, T.T. (2019). Semi-supervised inference: general theory and estimation of means. *Annals of Statistics 47, 2538–2566.*

Zhu, K. and Ling, S. (2015). LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises. *Journal of the American Statistical Association 110, 784–794.*

Zhu, K. (2016). Bootstrapping the portmanteau tests in weak auto-regressive moving average models. *Journal of the Royal Statistical Society Series B 78, 463–485.*