ANNALS OF ECONOMICS AND FINANCE **22-1**, 1–24 (2021)

Quantifying Diseconomies Of Scale For Mutual Funds

Ying Liao, Cuixia Li, Lei Jiang, and Liang Peng*

The fund size is highly persistent and correlated with risk factor loadings. Hence, it is unrealistic to assume constant diseconomies of scale over a long time. The traditional two-step method underestimates the uncertainty of diseconomies of scale. We propose a one-step procedure with a random weighted bootstrap method to infer diseconomies of scale using rolling windows, which effectively solves the problems. Our empirical analysis using actively-managed U.S. equity mutual funds supports diseconomies of scale, and simulations show that our rigorous method outperforms the two-step one in terms of precise estimating uncertainty.

Key Words: Diseconomies of scale; Fixed effects panel regression; Mutual funds.
JEL Classification Numbers: G23, C58.

1. INTRODUCTION

By assuming diseconomies of scale in actively-managed mutual funds, Berk and Green (2004) find that fund managers cannot outperform the factor benchmarks in the equilibrium. Since then, several papers use this crucial assumption in mutual fund researches; see Dangl, Wu and Zechner (2008), Stoughton, Wu and Zechner (2011), and Brown and Wu (2016). Therefore, it is essential to estimate and quantify diseconomies of scale using the sample of actively-managed U.S. equity mutual funds. Starting from Chen et al. (2004), researchers use a two-step procedure to estimate the marginal effect of the total net asset on fund future performance as the measure of diseconomies of scale. More specifically, Chen et al. (2004) first estimate fund alpha in time-series regressions based on popular benchmarks such as the one-factor model in Jensen (1968), the three-factor model in Fama and French (1993), and the four-factor model in Carhart (1997). The

* Liao: School of Statistics, Jiangxi University of Finance and Economics; Li: Corresponding author. School of Mathematics and Statistics, Xuzhou University of Technology. lcx@xzit.edu.cn; Jiang: Department of Finance, School of Economics and Management, Tsinghua University; Peng: Department of Risk Management and Insurance, Georgia State University.

1529-7373/2021

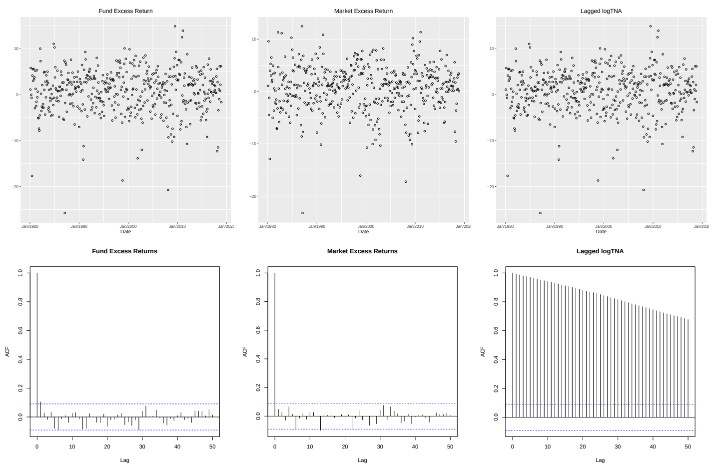
second step regresses alpha on the total net asset (TNA) of mutual funds in the previous month, together with other control variables such as fund flow and fund family size. They report the time-series averages of the slope coefficients in Fama and MacBeth (1973) regressions and the corresponding t-statistics adjusted for serial correlation using the method in Newey and West (1987) with lags of order three. Based on the signs of the coefficients and the corresponding t-statistics, they conclude that a fund with a higher TNA underperforms one with a smaller TNA significantly. Later, Edelen, Evans and Kadlec (2007), Christoffersen, Keim and Musto (2008), and Yan (2008) also find supporting evidence on diseconomies of scale by using a similar procedure. Recently, Pástor, Stambaugh and Taylor (2015) argue that, because of the unobservable skill of fund managers, there is an omitted variable bias in estimating the TNA's effect on fund future performance. They propose to employ fixed effects panel regression instead of Fama and MacBeth (1973) regressions in the second step and find no evidence of diseconomies of scale. The contradictory evidence about the strength of the diseconomy of scale in the literature motivates us to investigate the traditional two-step procedure carefully. Chen, Hribar and Melessa (2018) find that this two-step regression framework leads to incorrect inference in accounting research because of bias in estimating the coefficient and its standard error. In this paper, we notice two additional unique econometric properties of the variables in the study of diseconomies of scale in mutual funds, the highly persistent fund size and the correlation between risk factor loadings and the fund size.

Pollet and Wilson (2008) show that, when funds become bigger, fund managers tend to buy big stocks, which increases the average holding stock size of their funds. Busse, Chordia, Jiang and Tang (2019) find that to avoid high trade costs, funds with a large TNA tend to hold big stocks even after controlling fund styles. Both papers indicate that the total net asset of mutual funds is positively correlated with the stock size held by funds. By construction, big stocks have lower Small Minus Big (SMB) Betas. Therefore, we expect that big funds have smaller exposures to the SMB factor. Furthermore, Yan (2008) finds that big funds tend to hold stocks with a lower book to market ratio, making the High Minus Low (HML) Beta lower than that of small funds. Given this stylized evidence on the correlation between the fund size and factor loadings, when using the two-step procedure as in Chen et al. (2004), the estimation error for fund performance in the first step accumulates to the second step, leading to a severely biased inference of diseconomies of scale for mutual funds[1]. The simulation study in Section 4 well supports this argument.

[1] Pástor, Stambaugh and Taylor (2015) and Zhu (2018) directly subtract the Morningstar benchmark return from fund return without estimating Beta to calculate fund performance.

Furthermore, because skills of mutual fund managers cannot be observed and effectively controlled and are positively correlated with both fund size and fund performance, Pástor, Stambaugh and Taylor (2015) employ fixed effects panel regression to solve the omitted variable bias issue in the second step. Their t-statistics use the heteroskedasticity-robust variance estimation but ignore the uncertainty of estimating fund performance in the first step. Because most funds have a small sample size (ranging from tens to hundreds) compared to the larger number of available funds (more than 3000), their method may underestimate the variance. For fixed effects regression, Stock and Watson (2008) provide a bias corrected variance estimation, Cameron, Gelbach and Miller (2008) propose bootstrap methods for estimating the variance, Petersen (2009) surveys methods for computing the variance in finance, and Gonçalves (2011) studies the moving blocks bootstrap method. However, we can not apply these variance estimations to the two-step procedure in Chen et al. (2004) and Pástor, Stambaugh and Taylor (2015) because we have to take into account both the fixed effects in the second step and the uncertainty of obtaining benchmark-adjusted returns in the first step.

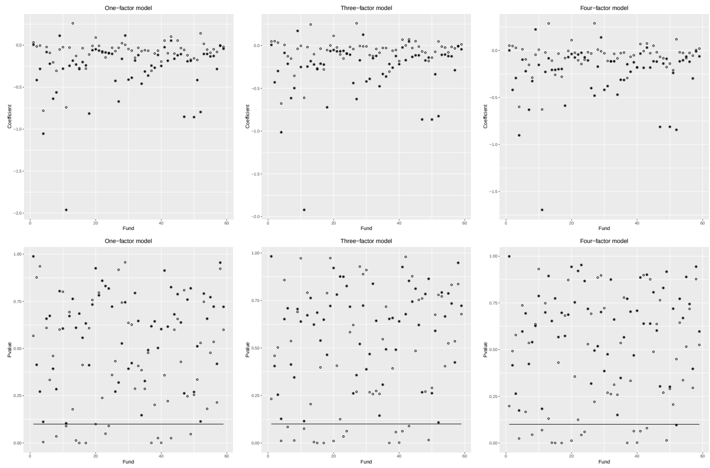
FIG. 1. *Scatter and autocorrelation plots.*



Using the fund (wficn=100019) with the maximal sample size in our dataset (467 observations), we plot the fund returns, market excess returns, and the lagged logTNAs in the upper panels, and the autocorrelation functions for each series in the bottom panels.

Finally, it is problematic to use long time-series and assume the constant marginal effect of fund size on performance because the commonly used fund size measure, logarithm of TNA (logTNA), is highly persistent. We illustrate the persistence using the fund with wficn = 100019, which

FIG. 2. *Estimates and p-values for testing zero coefficient of the lagged logTNA.*



For each fund with the maximal sample size (467 observations) in our dataset, we regress the fund returns by the market excess returns and the lagged logTNAs and plot the least squares estimates by circles and the instrumental variable estimates in Kostakis, Magdalinos and Stamatogiannis (2015) by stars in the upper panels and the p-values in the bottom panels, which are computed from the standard t-test (denoted by circles) and the instrumental variable test in Kostakis, Magdalinos and Stamatogiannis (2015) (denoted by stars) for testing zero coefficient of the lagged logTNA. The horizontal line is $y = 0.1$, indicating that points below the line reject the null hypothesis of zero coefficient at the 10% level.

starts from February 1980 and has the maximal sample size (467 observations) in our dataset. We plot its returns, the market excess returns and logTNAs, and their autocorrelation functions in Figure 1, which shows that the fund returns and market excess returns are stationary, but the logTNAs are nonstationary. This motivates Pástor, Stambaugh and Taylor (2015) and Zhu (2018) to employ an instrumental variable method to study the diseconomies of scale by treating the logTNAs as a unit root process. Unfortunately, the instrumental variable method in Pástor, Stambaugh and Taylor (2015) shows insignificant diseconomies of scale empirically, while Zhu (2018) shows that the diseconomies of scale are significant. Below we explain the problems of using instrumental variable methods, requiring a long time series, and the simulation setup in the two papers.

The model in the above two papers assumes that the fund return is linear to both the market excess return and lagged logTNA. Because the significant nonstationary predictor, the fund size, in a linear model can not be linearly correlated with the stationary fund return, the coefficient of the lagged logTNA must be zero if the model is well-specified. To confirm this, for each fund with the maximal sample size (467 observations) in our dataset, we regress the fund returns by the market excess returns and lagged logTNAs, estimate the coefficient of the lagged logTNA by the Least Squares Estimation (LSE) and the Instrumental Variable Estimation (IVX) in Kostakis, Magdalinos and Stamatogiannis (2015), and compute the p-values for testing zero coefficient of the lagged logTNA by the standard t-test and the IVX test in Kostakis, Magdalinos and Stamatogiannis (2015). From Figure 2, the IVX test shows insignificant diseconomies of scale except one fund under the four-factor model, and the t-test shows insignificance for most funds. The finding is in line with our argument that the regressor in a linear model can not be nonstationary if the dependent variable is stationary. Because the panel regression in Pástor, Stambaugh and Taylor (2015) and Zhu (2018) assumes that the coefficient of the lagged logTNA is independent of both funds and time, significant diseconomies of scale concluded from Zhu (2018) must hold for each fund, which contradicts Figure 2. In other words, when we use a long time series required by the IVX methods and a large number of funds in panel regression, it is unrealistic to assume that the coefficient of the lagged logTNA is constant. This assumption that the diseconomies of scale have a dynamic feature in our sample from 1980 to 2018 is consistent with the empirical observation in the mutual fund industry that both the number of mutual funds and the total net asset under management change dramatically over time with big increase after 2000 (Elton and Gruber, 2013). Pástor and Stambaugh (2012) and Pástor, Stambaugh and Taylor (2015) also find when the fund industry extends, it is hard for individual funds to generate outperformance, and small funds have less chance to outperform the large funds. However, the theory behind the instrumental variable methods in Pástor, Stambaugh and Taylor (2015) and Zhu (2018) requires that both the time-series sample size and the number of funds go to infinity. Furthermore, in the simulations, Pástor, Stambaugh and Taylor (2015) and Zhu (2018) regress the fund returns by the lagged logTNAs and regress the difference of logTNAs by the fund returns (with other fund characteristics as control variables). Because the second regression implies that the logTNAs at time t depend on the fund returns at time t, it is hard to believe the fund returns at time t only depend on the lagged logTNAs rather than the current logTNA at time t. When the logTNAs are a unit root process, the first regression assumes that fund returns are a unit root process, which implies from the second regression that the differences of logTNAs are a unit root process, i.e., logTNAs are not the unit root.

This paper uses more rigorous one-step fixed effects panel regression with a short time window such as three years, where the loadings of risk factors change with funds, but the coefficients of fund characteristics are independent of funds for measuring the marginal effect of TNA on fund performance. Hence, we overcome the econometric issues and can provide a dynamic picture of diseconomies of scale. This specification is similar to that in Chordia, Goyal and Shanken (2019) with individual stocks as test assets and in Busse, Jiang and Tang (2020) for quantifying how much fund factor loadings and characteristics can explain cross-sectional variation of mutual fund returns. Because we have to estimate some parameters and the fixed effects based on individual funds and other parameters based on all funds, it becomes nontrivial to estimate the asymptotic covariance of the estimators for the coefficients relating to fund characteristics. Using a short time series, the asymptotic theory for the employed inference should be valid when the number of funds goes to infinity. This paper uses a random weighted bootstrap method to quantify the uncertainty by allowing both fixed and divergent time-series sample size. Our simulation study shows that the one-step procedure significantly outperforms the traditional two-step procedure in terms of both mean squared error and coverage probability when diseconomies of scale do exist.

Applying the one-step procedure to the actively-managed U.S. equity mutual funds from January 1980 to December 2018 with a moving window of three years or five years, we find that the traditional two-step procedure underestimates the uncertainty of diseconomies of scale compared with the proposed one-step procedure, and both methods show significant diseconomies of scale consistent with Zhu (2018). To empirically justify the dynamics of diseconomies of scale, we use the t-test for two independent samples to test no difference in the coefficient of the lagged logTNA in the ith data window and the last data window. We notice that the independence assumption is problematic for two overlapping windows. Nevertheless, the p-values indicate that the null hypothesis of constant effect is rejected for most pairs, supporting that it is unrealistic to assume a constant coefficient of the lagged logTNA in panel regression when using long time series.

We organize this paper as follows. Section 2 describes the traditional two-step procedure and our one-step procedure for examining diseconomies of scale. Section 3 applies the one-step procedure to quantify diseconomies of scale in actively-managed U.S. equity mutual funds and compare the results with the two-step procedure. Section 4 uses simulated data to show that the one-step procedure outperforms the two-step procedure. Section 5 concludes. Theoretical derivations are put into the Appendix.

2. QUANTIFICATION OF DISECONOMIES OF SCALE

This section provides a detailed description of the traditional two-step procedure and our one-step procedure for estimating and quantifying diseconomies of scale in active mutual fund management.

For the ith fund at time t, $Y_{i,t}$ is the fund's excess return, and $\boldsymbol{X}_{i,t}$ is a d_1-dimensional control vector, including fund total net asset, fund family size, fund turnover, and others. In particular, we let the first element of $\boldsymbol{X}_{i,t}$ denote the logarithm of the fund's total net asset and the rest elements be any other explanatory variables. We use $\boldsymbol{F}_t$ to denote the d_2 factor benchmarks such as those in Jensen (1968), Fama and French (1993), and Carhart (1997). To study diseconomies of scale on fund performance, researchers such as Chen et al. (2004) first employ a factor model to get the benchmark-adjusted return (often called fund managers' skill) and then use this benchmark-adjusted return to fit fixed effect panel regression, which results in the following model:

$$Y_{i,t} = \alpha_i + \boldsymbol{\beta}_i^\tau \boldsymbol{F}_t + \varepsilon_{i,t} \text{ and } \alpha_i + \varepsilon_{i,t} = \mu_i + \boldsymbol{\gamma}^\tau \boldsymbol{X}_{i,t-1} + U_{i,t} \qquad (1)$$

for $t = 1 + a_i, \cdots, T_i + a_i$, and $i = 1, \cdots, n$, where $\varepsilon_{i,t}$ and $U_{i,t}$ are random errors with means zero, α_i is constant, but μ_i could be random. Here, we use a_i to allow funds with a different observation window and A^τ to denote the transpose of the matrix or vector of A. The interest is to infer $\boldsymbol{\gamma}$ and quantify the estimation uncertainty, especially for the coefficient of the fund total net asset, which is the measure of diseconomies of scale.

Under the above model, researchers often use a two-step inference procedure to estimate $\boldsymbol{\gamma}$. The first step estimates α_i and $\boldsymbol{\beta}_i$ by the least squares estimation based on data from the ith fund, giving estimators $\hat{\alpha}_i$, $\hat{\boldsymbol{\beta}}_i$, and $\hat{\varepsilon}_{i,t}$ for α_i, $\boldsymbol{\beta}_i$, and $\varepsilon_{i,t}$, respectively. The second step uses $\{\hat{\alpha}_i + \hat{\varepsilon}_{i,t} = Y_{i,t} - \hat{\boldsymbol{\beta}}_i^\tau \boldsymbol{F}_t : t = 1 + a_i, \cdots, T_i + a_i, i = 1, \cdots, n\}$ to fit fixed effects panel regression and get the estimator $\hat{\boldsymbol{\gamma}}$ for $\boldsymbol{\gamma}$, see Pástor, Stambaugh and Taylor (2015). Because big funds tend to hold big stocks (Pollet and Wilson, 2008 and Busse, Chordia, Jiang and Tang, 2019) and stock with low book to market ratio (Yan, 2008), it means that those funds have lower Small Minus Big (SMB) Betas and High Minus Low (HML) Beta. Therefore, this two-step inference is biased since $\boldsymbol{F}_t$ and $\boldsymbol{X}_{i,t-1}$ are correlated, and it is difficult to quantify the estimation uncertainty when we use monthly data, where the number of mutual funds is around 3000 and much larger than the sample size of most funds ranging from tens to hundreds.

This paper uses the following one-step fixed effects panel regression to investigate diseconomies of scale on fund performance:

$$Y_{i,t} = \alpha_i + \boldsymbol{\beta}_i^\tau \boldsymbol{F}_t + \boldsymbol{\gamma}^\tau \boldsymbol{X}_{i,t-1} + U_{i,t},\ t = 1 + a_i, \cdots, T_i + a_i,\ i = 1, \cdots, n, \quad (2)$$

which combines the two equations in (1). An important advantage is that inference based on (2) will consider the dependence between $\boldsymbol{F}_t$ and $\boldsymbol{X}_{i,t-1}$. Here, α_i could be random, and the interest is to infer $\boldsymbol{\gamma}$. Chordia, Goyal and Shanken (2019) use similar specifications in the stock market by controlling both factor exposures and stock characteristics in cross-sectional regression with individual stocks as test assets. Busse, Jiang and Tang (2020) use a similar specification to quantify the fraction of the cross-sectional variance of mutual fund returns explained by factor loadings and fund holdings. Greenaway-McGrevy, Han and Sul (2012) study the above model when we do not observe $\{\boldsymbol{F}_t\}$ and both $n \to \infty$ and $T_1 = \cdots = T_n = T \to \infty$. Because this model involves some parameters depending on i and some parameters independent of i, we first estimate α_i's and $\boldsymbol{\beta}_i$'s based on individual funds as a function of $\boldsymbol{\gamma}$. Then we estimate $\boldsymbol{\gamma}$ by using all funds.

Put $\boldsymbol{\theta}_i = (\alpha_i, \boldsymbol{\beta}_i^\tau)^\tau$ and $\bar{\boldsymbol{F}}_t = (1, \boldsymbol{F}_t^\tau)^\tau$. The least squares estimators solve the following score equations

$$\begin{cases} \sum_{t=1+a_i}^{T_i+a_i} \{Y_{i,t} - \boldsymbol{\gamma}^\tau \boldsymbol{X}_{i,t-1} - \boldsymbol{\theta}_i^\tau \bar{\boldsymbol{F}}_t\} \bar{\boldsymbol{F}}_t = 0 \text{ for } i = 1, \cdots, n, \\ \sum_{i=1}^n \sum_{t=1+a_i}^{T_i+a_i} \{Y_{i,t} - \boldsymbol{\gamma}^\tau \boldsymbol{X}_{i,t-1} - \boldsymbol{\theta}_i^\tau \bar{\boldsymbol{F}}_t\} \boldsymbol{X}_{i,t-1} = 0, \end{cases} \tag{3}$$

which gives $\hat{\boldsymbol{\gamma}} = \boldsymbol{Q}_n^{-1} \boldsymbol{S}_n$ and

$$\hat{\boldsymbol{\theta}}_i = (\sum_{t=1+a_i}^{T_i+a_i} \bar{\boldsymbol{F}}_t \bar{\boldsymbol{F}}_t^\tau)^{-1} \sum_{t=1+a_i}^{T_i+a_i} Y_{i,t} \bar{\boldsymbol{F}}_t - (\sum_{t=1+a_i}^{T_i+a_i} \bar{\boldsymbol{F}}_t \bar{\boldsymbol{F}}_t^\tau)^{-1} \sum_{t=1+a_i}^{T_i+a_i} \bar{\boldsymbol{F}}_t \boldsymbol{X}_{i,t-1}^\tau \hat{\boldsymbol{\gamma}}$$

for $i = 1, \cdots, n$, where $\boldsymbol{Q}_n = \sum_{i=1}^n \boldsymbol{Q}_{n,i}$, $\boldsymbol{S}_n = \sum_{i=1}^n \boldsymbol{S}_{n,i}$,

$$\boldsymbol{Q}_{n,i} = \sum_{t=1+a_i}^{T_i+a_i} \boldsymbol{X}_{i,t-1} \boldsymbol{X}_{i,t-1}^\tau - (\sum_{t=1+a_i}^{T_i+a_i} \boldsymbol{X}_{i,t-1} \bar{\boldsymbol{F}}_t^\tau)(\sum_{t=1+a_i}^{T_i+a_i} \bar{\boldsymbol{F}}_t \bar{\boldsymbol{F}}_t^\tau)^{-1}(\sum_{t=1+a_i}^{T_i+a_i} \bar{\boldsymbol{F}}_t \boldsymbol{X}_{i,t-1}^\tau),$$

and

$$\boldsymbol{S}_{n,i} = \sum_{t=1+a_i}^{T_i+a_i} Y_{i,t} \boldsymbol{X}_{i,t-1} - (\sum_{t=1+a_i}^{T_i+a_i} \boldsymbol{X}_{i,t-1} \bar{\boldsymbol{F}}_t^\tau)(\sum_{t=1+a_i}^{T_i+a_i} \bar{\boldsymbol{F}}_t \bar{\boldsymbol{F}}_t^\tau)^{-1}(\sum_{t=1+a_i}^{T_i+a_i} Y_{i,t} \bar{\boldsymbol{F}}_t).$$

Write

$$\begin{aligned} & \boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \\ = & \sum_{i=1}^n \{\sum_{t=1+a_i}^{T_i+a_i} U_{i,t} \boldsymbol{X}_{i,t-1} - (\sum_{t=1+a_i}^{T_i+a_i} \boldsymbol{X}_{i,t-1} \bar{\boldsymbol{F}}_t^\tau)(\sum_{t=1+a_i}^{T_i+a_i} \bar{\boldsymbol{F}}_t \bar{\boldsymbol{F}}_t^\tau)^{-1}(\sum_{t=1+a_i}^{T_i+a_i} U_{i,t} \bar{\boldsymbol{F}}_t)\} \\ = & \sum_{i=1}^n \sum_{t=1+a_i}^{T_i+a_i} U_{i,t} \{\boldsymbol{X}_{i,t-1} - (\sum_{s=1+a_i}^{T_i+a_i} \boldsymbol{X}_{i,s-1} \bar{\boldsymbol{F}}_s)(\sum_{s=1+a_i}^{T_i+a_i} \bar{\boldsymbol{F}}_s \bar{\boldsymbol{F}}_s^\tau)^{-1} \bar{\boldsymbol{F}}_t\}. \end{aligned}$$

Put $\boldsymbol{\Omega}_n = \sum_{i=1}^n \boldsymbol{\Omega}_{n,i}$, where $\boldsymbol{\Omega}_{n,i} = \sum_{t=1+a_i}^{T_i+a_i} E(U_{i,t}^2) \boldsymbol{W}_{i,t} \boldsymbol{W}_{i,t}^\tau$ and

$$\boldsymbol{W}_{i,t} = \boldsymbol{X}_{i,t-1} - (\sum_{s=1+a_i}^{T_i+a_i} \boldsymbol{X}_{i,s-1} \bar{\boldsymbol{F}}_s)(\sum_{s=1+a_i}^{T_i+a_i} \bar{\boldsymbol{F}}_s \bar{\boldsymbol{F}}_s^\tau)^{-1} \bar{\boldsymbol{F}}_t.$$

Under some regularity conditions (see Stock and Watson (2008) or Arrelano (2003)), there exists some $\delta > 0$ such that as $n \to \infty$,

$$\begin{cases} \boldsymbol{\Omega}_n / \sum_{i=1}^n T_i \xrightarrow{p} \Omega \text{ being nonsingular}, \ \sum_{i=1}^n E(||\boldsymbol{\Omega}_{n,i} / \sum_{j=1}^n T_j||^{1+\delta}) \to 0, \\ \boldsymbol{Q}_n / \sum_{i=1}^n T_i \xrightarrow{p} Q \text{ being nonsingular}, \ \sum_{i=1}^n E(||\boldsymbol{Q}_{n,i} / \sum_{j=1}^n T_j||^{1+\delta}) \to 0, \\ \sum_{i=1}^n E\{|| \sum_{t=1+a_i}^{T_i+a_i} U_{i,t} \boldsymbol{W}_{i,t} / \sum_{j=1}^n T_j||^{2+\delta}\} \to 0, \\ \frac{1}{\sqrt{\sum_{i=1}^n T_i}} \boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} N(0, \boldsymbol{\Omega}). \end{cases} \tag{4}$$

The conditions above allow some or all of T_i's to be finite. This is consistent with the empirical observation in mutual funds that the number of monthly returns for each fund ranges from tens to hundreds and is much smaller than the number of funds around 3000.

Note that we do not copy the existing regularity conditions for ensuring (4). Instead, we use (4) to prove that the proposed random weighted bootstrap method below is valid for quantifying the uncertainty of $\hat{\boldsymbol{\gamma}}$.

To quantify the uncertainty of $\hat{\boldsymbol{\gamma}}$, a heteroskedasticity-robust covariance estimator for $\boldsymbol{\Omega}_n$ is

$$\hat{\boldsymbol{\Omega}}_n = \sum_{i=1}^n \sum_{t=1+a_i}^{T_i+a_i} \hat{U}_{i,t}^2 \boldsymbol{W}_{i,t} \boldsymbol{W}_{i,t}^\tau,$$

where $\hat{U}_{i,t} = Y_{i,t} - \hat{\boldsymbol{\gamma}}^\tau \boldsymbol{X}_{i,t-1} - \hat{\boldsymbol{\theta}}_i^\tau \bar{\boldsymbol{F}}_t$ for $t = 1 + a_i, \cdots, T_i + a_i$ and $i = 1, \cdots, n$. When $\boldsymbol{F}_t = \boldsymbol{0}$ and $T_1 = \cdots = T_n = T$ is fixed, $(\hat{\boldsymbol{\Omega}}_n - \boldsymbol{\Omega}_n)/(nT)$ does not converge to zero in probability, and Stock and Watson (2008) propose a bias corrected covariance estimation. Because we use monthly fund returns to evaluate mutual fund performance, T_i's are much smaller than n, which makes the approximation error of $\hat{\boldsymbol{\Omega}}_n - \boldsymbol{\Omega}_n$ non-negligible. Instead of generalizing the idea in Stock and Watson (2008) and Cameron, Gelbach and Miller (2008) to our setting, where some parameters depend on individual funds, this paper uses a random weighted bootstrap procedure to quantify the uncertainty of $\hat{\boldsymbol{\gamma}}$. For dealing with heteroscedastic errors, Jing, Ying and Wei (2001) apply the random weighted bootstrap method to estimating equations, and Zhu (2016, 2019) apply the random weighted bootstrap procedure to time series models. Chiang, James and Wang (2005) and Zheng et al. (2018) argue that the random weighted bootstrap method is more computationally efficient than other bootstrap methods. Below we adopt the random weighted bootstrap procedure to our one-step panel regression.

Step 1) Draw a random sample δ_i^b for $i = 1, \cdots, n$ from a distribution with mean one and variance one, say a standard exponential distribution.

Step 2) Solve the following random weighted score equations

$$\begin{cases} \delta_i^b \sum_{t=1+a_i}^{T_i+a_i} \{Y_{i,t} - \boldsymbol{\gamma}^\tau \boldsymbol{X}_{i,t-1} - \boldsymbol{\theta}_i^\tau \bar{\boldsymbol{F}}_t\} \bar{\boldsymbol{F}}_t = 0 \text{ for } i = 1, \cdots, n, \\ \sum_{i=1}^n \delta_i^b \sum_{t=1+a_i}^{T_i+a_i} \{Y_{i,t} - \boldsymbol{\gamma}^\tau \boldsymbol{X}_{i,t-1} - \boldsymbol{\theta}_i^\tau \bar{\boldsymbol{F}}_t\} \boldsymbol{X}_{i,t-1} = 0. \end{cases}$$

Denote this resulted least squares estimator of $\boldsymbol{\gamma}$ by $\hat{\boldsymbol{\gamma}}^b$.

Step 3) Repeat the above two steps B times to get $\{\hat{\boldsymbol{\gamma}}^b\}_{b=1}^B$.

Using the above procedure, we estimate the asymptotic covariance of $\hat{\boldsymbol{\gamma}}$ by

$$\hat{\boldsymbol{\Sigma}}_{RWB} = \frac{1}{B} \sum_{b=1}^{B} (\hat{\boldsymbol{\gamma}}^b - \hat{\boldsymbol{\gamma}})(\hat{\boldsymbol{\gamma}}^b - \hat{\boldsymbol{\gamma}})^\tau = (\hat{\sigma}_{ij}).$$

Therefore, we can construct a two-sided confidence interval for γ_j of the jth element of $\boldsymbol{\gamma}$ with level $a \in (0,1)$ by using the normal approximation method, which gives

$$I_j^{NA}(a) = (\hat{\gamma}_j - z_{1-(1-a)/2}\hat{\sigma}_{jj},\ \hat{\gamma}_j + z_{1-(1-a)/2}\hat{\sigma}_{jj}),$$

where z_a denotes the a-th quantile of the standard normal distribution. When the distribution of $\hat{\gamma}_j$ is a bit away from a normal distribution, one can use the empirical distribution of the bootstrap sample. More specifically, put $\Delta_{j,b} = \hat{\gamma}_j^b - \hat{\gamma}_j$ for $b = 1, \cdots, B$, and let $\Delta_{j,B:1} \leq \cdots \leq \Delta_{j,B:B}$ denote the order statistics of $\Delta_{j,1}, \cdots, \Delta_{j,B}$. Then, we construct the two-sided confidence interval for γ_j at the level a as

$$I_j(a) = (\hat{\gamma}_j - \Delta_{j,B:[B(1+a)/2]}, \quad \hat{\gamma}_j - \Delta_{j,B:[B(1-a)/2]}).$$

THEOREM 1. *Suppose (4) hold. Then,*

$$\frac{1}{\sqrt{\sum_{i=1}^n T_i}} \boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}}^b - \hat{\boldsymbol{\gamma}}) \xrightarrow{d} N(0, \boldsymbol{\Omega}) \text{ as } n \to \infty.$$

Using (4), an application of standard asymptotic theory leads to

$$\frac{1}{\sqrt{\sum_{i=1}^n T_i}} \boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}}^b - \hat{\boldsymbol{\gamma}}) \xrightarrow{d} N(0, \boldsymbol{\Omega}) \text{ as } n \to \infty,$$

which ensures that the above variance estimation and confidence intervals are asymptotically correct. Again, we do not require $T_i \to \infty$ as $n \to \infty$.

3. EMPIRICAL ANALYSIS

We obtain the sample of actively-managed U.S. mutual funds from January 1980 to December 2018 from the Center for Research in Security Prices (CRSP) Survivorship Bias Free Mutual Fund Database. The variables of interest are monthly net returns and the total net asset (TNA) at the end of each month. To measure fund performance, we use Jensen (1968) market factor model, Fama and French (1993) three-factor model, and Carhart (1997) four-factor model to estimate alpha based on monthly fund returns. The factors, including CRSP value-weighted excess market return (Mktrf), size (SMB), book-to-market (HML), and momentum (UMD), are obtained from Ken French's website[2].

TABLE 1.

Summary statistics.

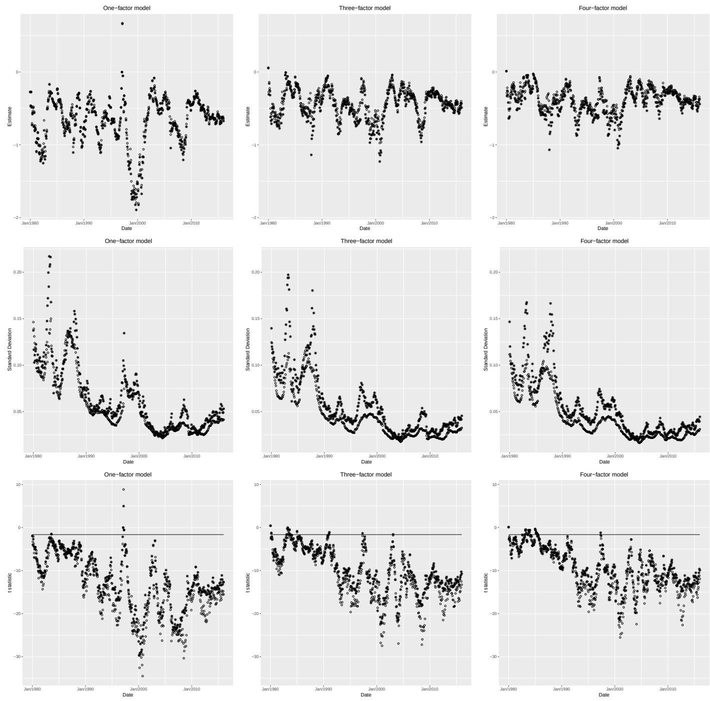
	All	Small	Group2	Group3	Group4	Big	Small-Big	t-statistic
NoF	1149	230	230	230	230	229		
Net	0.9235	0.9858	0.9358	0.8978	0.9051	0.8902	0.096	2.89
TNA	1.0651	0.0379	0.1064	0.2580	0.6509	4.2803	-4.242	-32.64
Age	17.2727	11.3945	14.5323	15.6644	19.0528	25.7563	-14.362	-121.97
FamTNA	98.5892	36.1972	52.2037	85.1416	115.6269	204.0239	167.827	-18.77
Turnover	80.5873	94.8255	86.0332	83.5474	76.0341	62.4041	32.421	43.26
Flow	0.6575	1.7573	0.5904	0.4904	0.3176	0.1251	1.632	20.08
Expense	1.1436	1.3390	1.2248	1.1712	1.0700	0.9166	0.427	104.24

For all actively-managed U.S. equity funds with at least 24 valid observations from January 1980 to December 2018, we sort all funds by TNA into five groups each month, report the time-series average number of funds in each portfolio, the time-series averages of the monthly cross-sectional means for the fund characteristics in each portfolio, and the difference in means between the two extreme portfolios including the total net asset ($ billion), fund age (in years), fund family size ($ billion), net return, turnover, flow, and expense ratio (in percentage point). These variables are defined in Section 3.

Fund characteristics for each share class, such as fund expense ratio and turnover, are from CRSP too. We define fund flow as the monthly net growth in fund assets beyond capital gains and dividends. We aggregate variables in different share classes based on the unique identifier of "wficn" in "MFLINK1" provided by Wharton Research Data Services. The fund-level total net asset is the sum of TNA across different share classes of the fund. The net return, expense ratio, turnover ratio, and flow are the TNA-weighted averages of them across all fund share classes, respectively. Fund age is the age of the oldest share class in the fund. Fund family size is the TNA summation of each fund in a fund family (excluding the fund itself). We assume that funds with the same management company name belong to the same fund family. These variables are commonly used in the

[2]https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/datalibrary.html

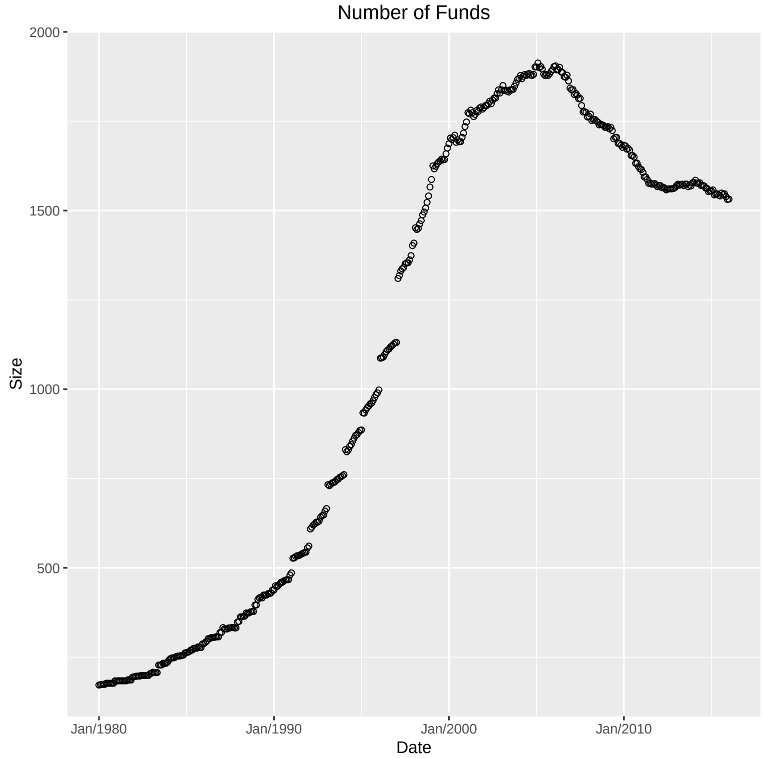
FIG. 3. *Estimates, standard deviations, and t-statistics based on three-year data.*



Using every three-year data, we plot the estimates (upper panels), standard deviations (middle panels), and t-statistics (bottom panels) for testing zero coefficient of the lagged logTNA computed from the two-step method (marked by circles) and from the one-step method (marked by stars). The horizontal line is $y = -1.645$, indicating that points below this line reject the null hypothesis of zero coefficient at the 10% level.

scale and performance literature. We winsorize the expense ratio, turnover ratio, and flow at 0.5% to avoid extreme values. In the regressions, we take the logarithm for TNA, turnover, fund age, and fund family size. We impose the filters to narrow our sample down to the actively-managed U.S. equity mutual funds based on Kacperczyk et al. (2008). Furthermore, we also exclude funds with an average percentage of common stocks lower than 80% of the total net asset. We identify index funds, ETF, and other passive funds using their names and the CRSP index fund identifier following Busse and Tong (2012) and Ferson and Lin (2014). We exclude funds with the following Investment Objective Codes in the Thomson Reuters Mutual Fund Holdings database: International, Municipal Bonds, Bond and Preferred, Balanced, and Metals. Following Elton et al. (2001), we exclude funds with less than \$15 million in TNA, and we address incubation bias

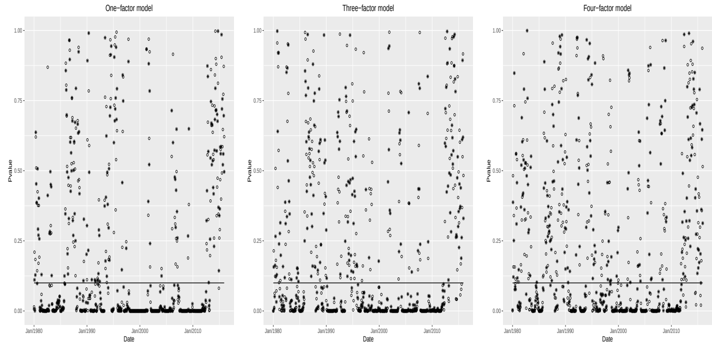
FIG. 4. The number of funds in each of three-year data windows.



We plot the number of actively-managed U.S. funds in our sample for each of the three-year data windows.

following Evans (2010). We restrict funds with a minimal sample size 24 and fit the model (2). In Table 1, for each month, we sort funds in our sample by the total net asset of the previous month into quintiles. Quintile 1 contains funds with the lowest TNA (Small), and Quintile 5 contains funds with the highest TNA (Big). We report the average monthly net returns, fund sizes, fund age (in years), fund family size, turnover, flow, and expense ratio in each portfolio. The corresponding t-statistic to test the difference in means is provided for each characteristic. First of all, the average fund size is about 1 billion USD in our sample. Also, the average total net asset of funds in the smallest quantile is 38 million USD with the equal-weighed average of net return of 0.99% a month. In contrast, the net return of the largest 20% funds is 0.89%. We could see a clear decreasing trend of returns from small funds to big funds, with a significant positive difference in returns between the two extreme TNA portfolios (0.10%). Furthermore,

FIG. 5. P-values for testing the difference based on the ith three-year data window and the last one to be zero.



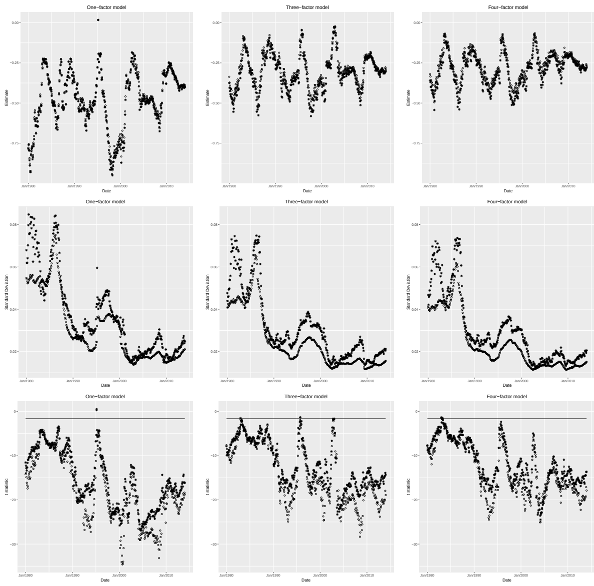
Using estimates in the ith three-year data window and the last three-year data window, we plot the p-values of the t-test for two independent samples computed from the two-step method (marked by circles) and from the one-step method (marked by stars). The horizontal line is $y = 0.1$, indicating that points below this line reject the null hypothesis of no difference in the coefficient of the lagged logTNA at the 10% level.

the table shows that small funds are younger in smaller fund families with higher turnover, higher inflow, and larger expense ratios, which all match the results from Chen et al. (2004).

As argued in the introduction, the persistence of the logTNA makes it unrealistic to assume a constant coefficient of the lagged logTNA in panel regression if we use long time series. Hence, we apply the panel regression to data with a moving window of three years and five years. To estimate and quantify diseconomies of scale, we consider the one-factor, three-factor, and four-factor models, and calculate estimates, standard deviations, and t-statistics by using our one-step inference and the random weighted bootstrap method with $B = 5000$ repetitions. For a comparison purpose, we also calculate these quantities using the two-step approach in Pástor, Stambaugh and Taylor (2015) with the heteroskedasticity-robust covariance matrix estimation in Stock and Watson (2008). The t-statistic is the ratio of estimate to its standard deviation. A calculation of the variance inflation factor (VIF) shows that severe multicollinearity exists between total net asset of funds and other fund characteristics, including fund age, fund family size, turnover, flow, and expense ratio. For example, the VIFs based on the fund with wficn=100019 are $28.99, 1.78, 2.96, 2.42, 263.16$, and 156.25 for the logTNA, expense ratio, log of turnover ratio, fund flow, log of fund age, and log of fund family size, respectively. Therefore, due to the multicollinearity issue, we use logTNA as the only independent variable in the panel regression.

Using every three-year data, we plot the estimates, standard deviations, and t-statistics for our one step procedure (marked by stars) and the two-

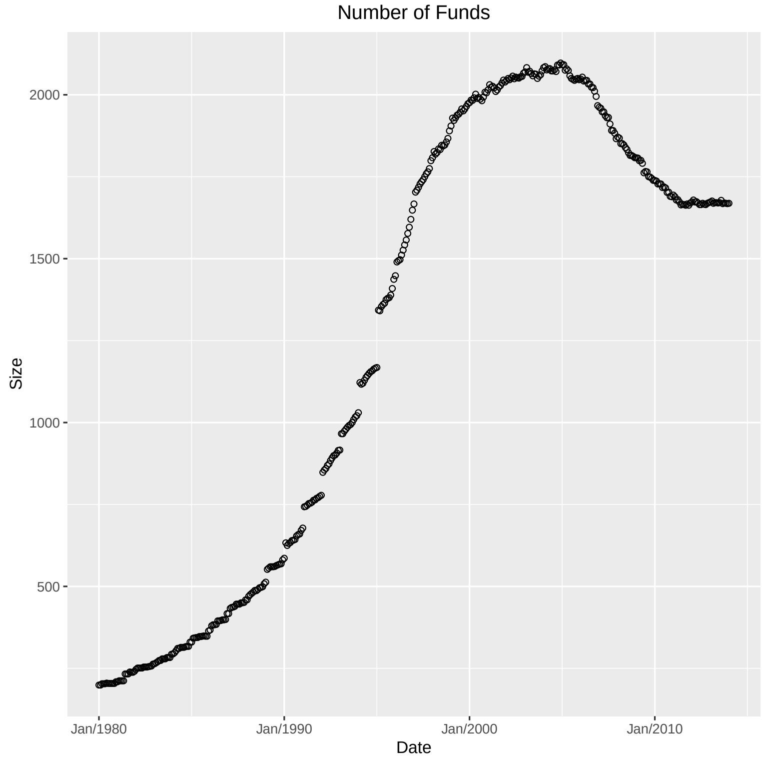
FIG. 6. Estimates, standard deviations, and t-statistics based on five-year data.



Using every five-year data, we plot the estimates (upper panels), standard deviations (middle panels), and t-statistics (bottom panels) for testing zero coefficient of the lagged logTNA computed from the two-step method (marked by circles) and from the one-step method (marked by stars). The horizontal line is $y = -1.645$, indicating that points below this line reject the null hypothesis of zero coefficient at the 10% level.

step approach in Pástor, Stambaugh and Taylor (2015) (marked by circles) in Figure 3, which shows that the estimates of the two approaches are similar, the standard deviations from the one-step procedure are bigger than those from the two-step method, and the absolute t-statistics from the two-step procedure are larger than those from the one-step procedure. Hence, the two-step procedure underestimates the asymptotic variance compared with the one-step procedure. Overall, we conclude that, although both methods confirm the diseconomies of scale over the sample from 1980 to 2018, our method is more precise in estimating uncertainty and the strength of diseconomies of scale. We also plot the number of funds in each of the three-year moving windows in Figure 4. To confirm the dynamic feature of diseconomies of scale, we check whether the t-statistics in two different three-year data windows are significantly different. We use the t-test for two independent samples, i.e., the difference between the estimates in the ith three-year data window and the last three-year data window divided by the squared root of the sum of their variances. We notice that the independence assumption is problematic for two overlapping windows.

FIG. 7. The number of funds in each of five-year data windows.

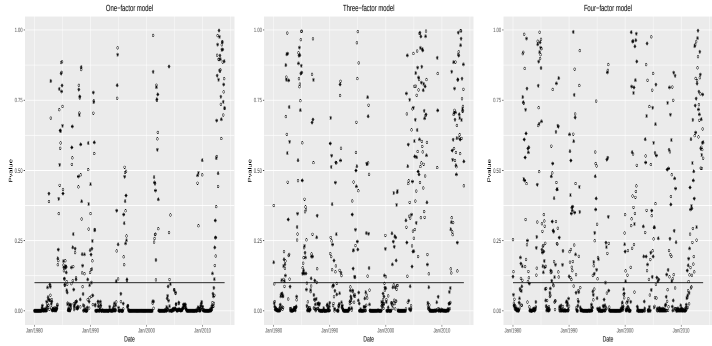


We plot the number of actively-managed U.S. funds in our sample for each of the five-year data windows.

Nevertheless, Figure 5 indicates that most of the pairs are significantly different.

Using every five-year data, we plot the estimates, standard deviations, and t-statistics in Figure 6. Figure 7 plots the number of funds in each of the five-year moving windows, and Figure 8 plots the p-values of the t-test for testing no difference in the coefficient of the lagged logTNA between the ith five-year data window and the last five-year data window. Again, we conclude that the two-step procedure underestimates the asymptotic variance compared with the one-step procedure, and it is unrealistic to assume a constant coefficient of the lagged logTNA in panel regression.

FIG. 8. p-values for testing no difference based on two five-year data windows.



Using estimates in the ith three-year data window and the last three-year data window, we plot the p-values of the t-test for two independent samples computed from the two-step method (marked by circles) and from the one-step method (marked by stars). The horizontal line is $y = 0.1$, indicating that points below this line reject the null hypothesis of no difference in the coefficient of the lagged logTNA at the 10% level.

4. SIMULATION STUDY

This section investigates the finite sample performance of the employed one-step procedure to ensure that conclusions made for our mutual fund data above are sound. For a comparison purpose, we also implement the two-step approach in Pástor, Stambaugh and Taylor (2015) with the heteroskedasticity-robust covariance matrix estimation in Stock and Watson (2008), which ignores the uncertainty in the first step.

We draw random samples from the following panel regression with 1000 repetitions:

$$Y_{i,t} = \alpha_i + \gamma X_{i,t-1} + \boldsymbol{\beta}_i^\tau \boldsymbol{F}_t + U_{i,t},\ t = 1, \cdots, T_i,\ i = 1, \cdots, n, \quad (5)$$

where $n = 3073$ is the total number of funds, and T_i is the sample size of the ith fund in our real dataset analyzed in Section 4 below. Figure 9 plots the sample sizes for each of these 3073 funds ranging from tens to hundreds.

To have a setting close to our real dataset, we fit the above model to the real dataset with $\boldsymbol{F}_t = (F_{t,1}, \cdots, F_{t,4})^\tau$ being the benchmarks in the four-factor model and $X_{i,t}$ being the logarithm of the total net asset. Figure 10 plots the least squares estimates for α_i and $\boldsymbol{\beta}_i = (\beta_{i,1}, \cdots, \beta_{i,4})^\tau$.

Further, using the data for the first fund, we estimate the means and covariance of $\{(X_{1,t}, F_{t,1}, \cdots, F_{t,4})^\tau\}_{t=1}^{T_1}$ and the mean and variance of residuals $\{\hat{U}_{1,t}\}_{t=1}^{T_1}$. Using these estimates, we independently generate random samples for $\{(X_t, F_{t,1}, \cdots, F_{t,4})^\tau\}_{t=1}^{\max(T_1,\cdots,T_n)}$ from the multivariate nor-

FIG. 9. Sample sizes of the 3073 mutual funds from January 1980 to December 2018.

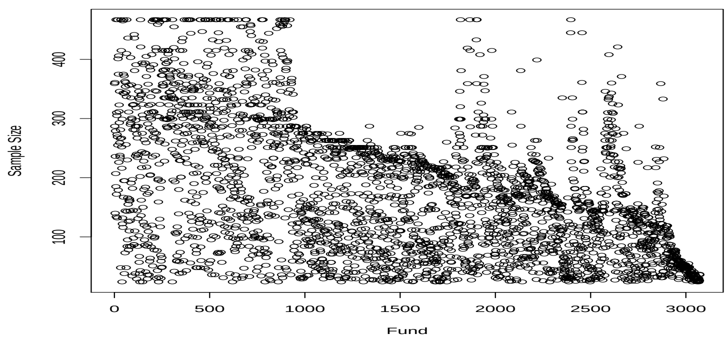
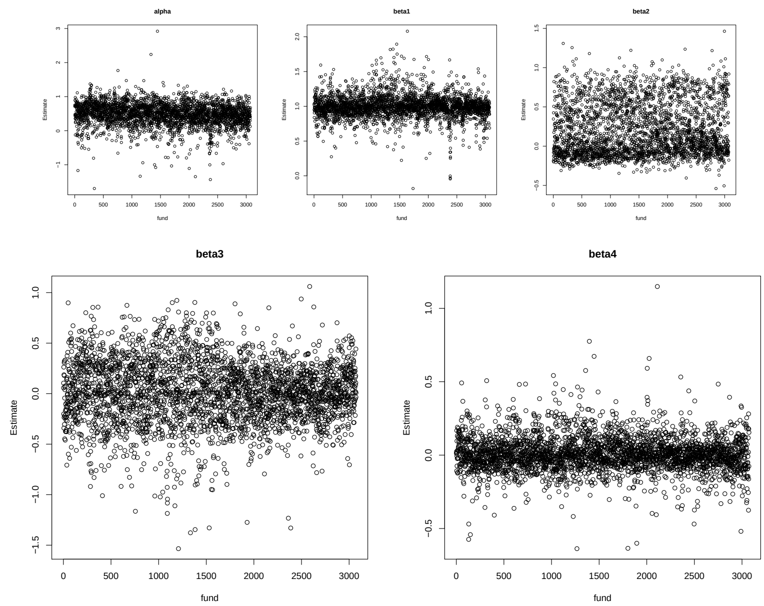


FIG. 10. Least squares estimates for α_i and $\boldsymbol{\beta}_i = (\beta_{i,1}, \cdots, \beta_{i,4})^\tau$ based on the 3073 mutual funds from January 1980 to December 2018.



mal distribution

$$N\left(\begin{pmatrix}7.312\\0.633\\0.097\\0.153\\0.577\end{pmatrix}, \begin{pmatrix}1.033 & -0.232 & 0.175 & 0.193 & -0.108\\ -0.232 & 17.682 & 2.915 & -2.182 & -4.588\\ 0.175 & 2.915 & 10.002 & -2.377 & 0.712\\ 0.913 & -2.182 & -2.377 & 8.596 & -2.689\\ -0.108 & -4.588 & 0.712 & -2.689 & 21.816\end{pmatrix}\right),$$

for $\{U_{i,t} : t = 1, \cdots, T_i, i = 1, \cdots, n\}$ from $N(0, 0.984^2)$, and set $X_{i,t} = Z_{i,t}X_t$, where $Z_{i,t}$'s for $t = 1, \cdots, T_i, i = 1, \cdots, n$ are independent random variables with the standard exponential distribution function. Finally, we generate $Y_{i,t}$'s from the model (5) with $\gamma = -0.4, -0.3, -0.2, -0.1$, and 0.

TABLE 2.

Simulation results for the unbalanced panel regression.

One-step	γ	Mean	SD	RMSE	Estimated SD	$I_1^{NA}(0.9)$	$I_1^{NA}(0.95)$
	-0.4	-0.400007	0.000187	0.000187	0.000183	0.895	0.937
	-0.3	-0.300007	0.000187	0.000187	0.000183	0.895	0.937
	-0.2	-0.200007	0.000187	0.000187	0.000183	0.895	0.937
	-0.1	-0.100007	0.000187	0.000187	0.000183	0.895	0.937
	0	-0.000007	0.000187	0.000187	0.000183	0.895	0.937
Two-step	γ	Mean	SD	RMSE	Estimated SD	$I_1^{NA}(0.9)$	$I_1^{NA}(0.95)$
	-0.4	-0.390683	0.000293	0.009322	0.000202	0	0
	-0.3	-0.293014	0.000250	0.006991	0.000192	0	0
	-0.2	-0.195345	0.000215	0.004660	0.000184	0	0
	-0.1	-0.097676	0.000191	0.002332	0.000180	0	0
	0	-0.000007	0.000183	0.000183	0.000178	0.895	0.937

We report the mean, standard deviation, and root mean squared error of $\hat{\gamma}$ and the average of the estimated standard deviations by the random weighted bootstrap method in the upper panel. We also compute these quantities by using the two-step inference in Pástor, Stambaugh and Taylor (2015) and the heteroskedasticity-robust covariance matrix estimation in Stock and Watson (2008), respectively, in the lower panel. We compute the coverage probabilities of the one-step and two-step approaches based on the normal approximation method.

Based on these simulated data, we compute our estimator $\hat{\gamma}$ and the standard deviation estimator by the random weighted bootstrap method with $B = 10000$. We also calculate the two-step estimator for γ in Pástor, Stambaugh and Taylor (2015) and the heteroskedasticity-robust covariance matrix estimation in Stock and Watson (2008). The t-statistic is the ratio of the estimator to its standard deviation. Using these 1000 repetitions, Table 2 reports the means, standard deviations, and root mean squared errors of these two estimators, the averages of these two standard deviation estimators, and the coverage probabilities of the normal approximation confidence intervals based on the t-statistics. As the distribution of $\hat{\gamma} - \gamma$ is independent of γ, the quantities for the one-step approach are the same except the average of $\hat{\gamma}$. Results in Table 2 show that the employed one-step procedure works very well in terms of both point estimation and interval estimation. In contrast, the two-step approach in Pástor, Stambaugh and Taylor (2015) has a significant bias for $\gamma \neq 0$ by comparing the simulated standard deviation with the root mean squared error and becomes worse as $|\gamma|$ is larger. Moreover, the heteroskedasticity-robust covariance matrix estimation in Stock and Watson (2008) severely underestimates the asymptotic variance, leading to zero coverage probability for the intervals constructed from the two-step approach.

TABLE 3.

Simulation results for the balanced panel regression with $n = 3073$.

One-step	γ	Mean	SD	RMSE	Estimated SD	$I_1^{NA}(0.9)$	$I_1^{NA}(0.95)$
	-0.4	-0.400006	0.000321	0.000321	0.000321	0.903	0.945
	-0.3	-0.300006	0.000321	0.000321	0.000321	0.903	0.945
	-0.2	-0.200006	0.000321	0.000321	0.000321	0.903	0.945
	-0.1	-0.200006	0.000321	0.000321	0.000321	0.903	0.945
	0	-0.000006	0.000321	0.000321	0.000321	0.903	0.945
Two-step	γ	Mean	SD	RMSE	Estimated SD	$I_1^{NA}(0.9)$	$I_1^{NA}(0.95)$
	-0.4	-0.372798	0.000880	0.027216	0.000398	0	0
	-0.3	-0.279600	0.000688	0.020412	0.000358	0	0
	-0.2	-0.186402	0.000510	0.013608	0.000325	0	0
	-0.1	-0.093204	0.000363	0.006806	0.000304	0	0
	0	-0.000006	0.000299	0.000299	0.000297	0.900	0.942

We set the number of time-series observations to be 60 for all funds and the number of funds to be 3073, report the mean, standard deviation, and root mean squared error of $\hat{\gamma}$ and the average of the estimated standard deviations by the random weighted bootstrap method in the upper panel. We also compute these quantities by using the two-step inference in Pástor, Stambaugh and Taylor (2015) and the heteroskedasticity-robust covariance matrix estimation in Stock and Watson (2008), respectively, in the lower panel. We compute the coverage probabilities of the one-step and two-step approaches based on the normal approximation method.

Next, we study balanced panel regression with $n = 3073$ and $n = 300$ but fixing the number of time-series observations at 60 for all funds. The rest of the settings is the same as above. Like Table 2, we calculate and report the estimates, standard deviations, and coverage probabilities in Table 3 for $n = 3703$ and Table 4 for $n = 300$, which show that the employed one-step procedure performs very well, but the two-step approach has a significant bias and zero coverage probability for $\gamma \neq 0$, and it is worse than the corresponding results in Table 2, which has a larger sample size for many funds.

In summary, the used one-step approach with the random weighted bootstrap method works very well for panel data regardless of the time series being longer or shorter. Without taking into account the dependence between predictors in the first step and the second step, the two-step approach in Pástor, Stambaugh and Taylor (2015) is biased, especially for panel data with shorter periods, and ignoring the uncertainty in the first step leads to very inaccurate confidence intervals.

TABLE 4.

Simulation results for the balanced panel regression with $n = 300$.

One-step	γ	Mean	SD	RMSE	Estimated SD	$I_1^{NA}(0.9)$	$I_1^{NA}(0.95)$
	-0.4	-0.400010	0.000974	0.000974	0.001026	0.878	0.933
	-0.3	-0.300010	0.000974	0.000974	0.001026	0.878	0.933
	-0.2	-0.200010	0.000974	0.000974	0.001026	0.878	0.933
	-0.1	-0.100010	0.000974	0.000974	0.001026	0.878	0.933
	0	-0.000010	0.000974	0.000974	0.001026	0.878	0.933
Two-step	γ	Mean	SD	RMSE	Estimated SD	$I_1^{NA}(0.9)$	$I_1^{NA}(0.95)$
	-0.4	-0.372804	0.001623	0.027245	0.001275	0	0
	-0.3	-0.279605	0.001355	0.020440	0.001145	0	0
	-0.2	-0.186407	0.001126	0.013640	0.001042	0	0
	-0.1	-0.093208	0.000965	0.006860	0.000975	0	0
	0	-0.000009	0.000908	0.000908	0.000952	0.911	0.961

We set the number of time-series observations to be 60 for all funds and the number of funds to be 300, report the mean, standard deviation, and root mean squared error of $\hat{\gamma}$ and the average of the estimated standard deviations by the random weighted bootstrap method in the upper panel. We also compute these quantities by using the two-step inference in Pástor, Stambaugh and Taylor (2015) and the heteroskedasticity-robust covariance matrix estimation in Stock and Watson (2008), respectively, in the lower panel. We compute the coverage probabilities of the one-step and two-step approaches based on the normal approximation method.

5. CONCLUSIONS

Whether large funds significantly underperform the small ones has received considerable attention in the mutual fund industry. In this paper, we argue that the two-step procedure in Chen et al. (2004) and Pástor, Stambaugh and Taylor (2015) underestimates the estimation uncertainty for diseconomies of scale. Further, the instrumental variable methods in Pástor, Stambaugh and Taylor (2015) and Zhu (2018) require a long time series theoretically and make it unrealistic to assume constant coefficient of the lagged logTNA in panel regression. To solve these problems, we use a one-step model and inference with a random weighted bootstrap method to fit fixed effects panel regression for a short time window. A simulation study shows that the employed one-step procedure outperforms the traditional two-step procedure in terms of interval estimation. An application to actively-managed U.S. equity mutual funds from 1980 to 2018 with a moving window of three-year or five-year data shows that the traditional two-step approach underestimates the standard deviation compared with the one-step procedure. Both methods find significant diseconomies of scale. We find that the strength of diseconomies of scale in different data windows varies significantly in most cases, which supports the time-dynamic feature of diseconomies of scale.

ACKNOWLEDGMENTS

We thank a reviewer for his/her helpful comments. Li's research was partly supported by the NSFC 11301236. Peng's research was partly supported by the Simons Foundation and the NSF grant of DMS-2012448.

APPENDIX: PROOF OF THEOREM 1

Put

$$\boldsymbol{Q}_n^b = \sum_{i=1}^n \delta_i^b \sum_{t=1+a_i}^{T_i+a_i} \boldsymbol{W}_{i,t}\boldsymbol{X}_{i,t-1}^\tau,$$

and write

$$\begin{aligned} & \boldsymbol{Q}_n(\hat{\gamma}^b - \hat{\gamma}) \\ = & \boldsymbol{Q}_n(\boldsymbol{Q}_n^b)^{-1}\sum_{i=1}^n(\delta_i^b - 1)\sum_{t=1+a_i}^{T_i+a_i} U_{i,t}\boldsymbol{W}_{i,t} \\ & +(\boldsymbol{Q}_n(\boldsymbol{Q}_n^b)^{-1} - I_{d_1\times d_1})\sum_{i=1}^n\sum_{t=1+a_i}^{T_i+a_i} U_{i,t}\boldsymbol{W}_{i,t}, \end{aligned} \tag{A.1}$$

where $I_{d_1\times d_1}$ denotes the $d_1 \times d_1$ identity matrix. By

$$\boldsymbol{Q}_n^b - \boldsymbol{Q}_n = \sum_{i=1}^n(\delta_i^b - 1)\sum_{t=1+a_i}^{T_i+a_i} \boldsymbol{W}_{i,t}\boldsymbol{X}_{i,t-1}^\tau$$

and (4), an application of the law of large numbers yields that as $n \to \infty$

$$\frac{\boldsymbol{Q}_n^b - \boldsymbol{Q}_n}{\sum_{i=1}^n T_i} \xrightarrow{p} \boldsymbol{0}, \text{ i.e., } \frac{\boldsymbol{Q}_n^b}{\sum_{i=1}^n T_i} \xrightarrow{p} \boldsymbol{Q}. \tag{A.2}$$

By (4) and the central limit theorem, we have

$$\frac{1}{\sqrt{\sum_{i=1}^n T_i}}\sum_{i=1}^n(\delta_i^b - 1)\sum_{t=1+a_i}^{T_i+a_i} U_{i,t}\boldsymbol{W}_{i,t} \xrightarrow{d} N(0, \boldsymbol{\Omega}) \text{ as } n \to \infty. \tag{A.3}$$

Hence, the theorem follows from equations (A.1)–(A.3).

REFERENCES

Arrelano, M., 2003. Panel Data Econometrics. *Oxford University Press.*

Berk, J. and R. Green, 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy* **112**, 1269-1295.

Brown, D. and Y. Wu, 2016. Mutual fund flows and cross-fund learning within families. *Journal of Finance* **71**, 383-424.

Busse, J. A., T. Chordia, L. Jiang, and Y. Tang, 2019. Transaction Costs, portfolio characteristics, and mutual fund performance. *Management Science.* Forthcoming.

Busse, J. A., L. Jiang, and Y. Tang, 2020. Double adjusted mutual fund performance. *Review of Asset Pricing Studies.* Forthcoming.

Busse, J. A. and Q. Tong, 2012. Mutual fund industry selection and persistence. *Review of Asset Pricing Studies* **2**, 245-274.

Cameron, A. C., J. B. Gelbach, and D. L. Miller, 2008. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics* **90**, 414-427.

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

Chen, J., H. Hong, M. Huang, and J. Kubik, 2004. Does fund size erode mutual fund performance? The role of liquidity and organization. *American Economic Review* **94**, 1276-1302.

Chen, W., P. Hribar, and S. Melessa, 2018. Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research* **56**, 751-796.

Chiang, C. T., L. F. James, and M. C. Wang, 2005. Random weighted bootstrap method for recurrent events with informative censoring. *Lifetime Data Analysis* **11**, 489-509.

Chordia, T., A. Goyal, and J. Shanken, 2019. Cross-sectional asset pricing with individual stocks: betas versus characteristics. *Working Paper, Emory University.*

Christoffersen, S., D. Keim, and D. Musto, 2008. Valuable information and costly liquidity: evidence from individual mutual fund trades. *Working Paper, University of Pennsylvania.*

Dangl, T., Y. Wu, and J. Zechner, 2008. Market discipline and internal governance in the mutual fund industry. *Review of Financial Studies* **21**, 2307-2343.

Edelen, R., R. Evans, and G. Kadlec, 2007. Scale effects in mutual fund performance: the role of trading costs. *Working Paper, UC Davis.*

Elton, E., and M. Gruber, 2013. Mutual funds. *In Handbook of the Economics of Finance* **2**, 1011-1061. Elsevier.

Elton, E., M. Gruber, and C. Blake, 2001. A first look at the accuracy of the CRSP mutual database and a comparison of the CRSP and Morningstar mutual fund databases. *Journal of Finance* **56**, 2415-2430.

Evans, R. B., 2010. Mutual fund incubation. *Journal of Finance* **65**, 1581-1611.

Fama, E. F. and K. R. French, 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33**, 3–56.

Fama, E. F. and J. D. MacBeth, 1973. Risk, return, and equilibrium: empirical tests. *Journal of Political Economy* **81**, 607-636.

Ferson, W. and J. Lin, 2014. Alpha and performance measurement: the effects of investor disagreement and heterogeneity. *Journal of Finance* **69**, 1565-1596.

Gonçalves, S., 2011, The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Econometric Theory 27, 1048-1082.*

Greenaway-McGrevy, R., C. Han, and D. Sul, 2012. Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics* **169**, 48-53.

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

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

Kacperczyk, M., C. Sialm, and L. Zheng, 2008. Unobserved actions of mutual funds. *Review of Financial Studies* **21**, 2379-2416.

Kostakis, A., T. Magdalinos, and M. Stamatogiannis, 2015. Robust econometric inference for stock return predictability. *Review of Financial Studies* **28**, 1506-1553.

Newey, W. K. and K. D. West, 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703-708.

Pástor, L. and R. F. Stambaugh, 2012. On the size of the active management industry. *Journal of Political Economy* **120**, 740-781.

Pástor, L., R. F. Stambaugh, and L. A. Taylor, 2015. Scale and skill in active management. *Journal of Financial Economics* **116**, 23-45.

Petersen, M., 2009. Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies* **22**, 435-480.

Pollet, J. and M. Wilson, 2008. How does size affect mutual fund behavior? *Journal of Finance* **63**, 2941-69.

Stock, J. H. and N. Watson, 2008. Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica* **76**, 155-174.

Stoughton, N., Y. Wu, and J. Zechner, 2011. Intermediated investment management. *Journal of Finance* **66**, 947-980.

Yan, X., 2008. Liquidity, investment style, and the relation between fund size and fund performance. *Journal of Financial and Quantitative Analysis* **43**, 741-767.

Zheng, Y., Q. Zhu, G. Li, and Z. Xiao, 2018. Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *Journal of the Royal Statistical Society Series B* **80**, 975-993.

Zhu, K., 2016. Bootstrapping the portmanteau tests in weak autoregressive moving average models. *Journal of the Royal Statistical Society B* **78**, 463-485.

Zhu, K., 2019. Statistical inference for autoregressive models under heteroscedasticity unknown form *Annals of Statistics* **6**, 3185–3215.

Zhu, M., 2018. Informative fund size, managerial kill, and investor rationality. *Journal of Financial Economics* **130**, 114-134.