International Review of Economics and Finance 86 (2023) 364–379

Contents lists available at ScienceDirect

International Review of Economics and Finance

journal homepage: www.elsevier.com/locate/iref

Improved inference in financial factor models

Elliot Beck [a,b], Gianluca De Nard [c,d,e,*], Michael Wolf [c]

[a] *Department of Banking and Finance, University of Zurich, 8032 Zurich, Switzerland*
[b] *Swiss National Bank, 8001 Zurich, Switzerland*
[c] *Department of Economics, University of Zurich, 8032 Zurich, Switzerland*
[d] *NYU Stern Volatility and Risk Institute, USA*
[e] *OLZ AG, Switzerland*

ARTICLE INFO

JEL classification:
codes
C12
C13
C21

ABSTRACT

Conditional heteroskedasticity of the error terms is a common occurrence in financial factor models, such as the CAPM and Fama–French factor models. This feature necessitates the use of heteroskedasticity consistent (HC) standard errors to make valid inference for regression coefficients. In this paper, we show that using weighted least squares (WLS) or adaptive least squares (ALS) to estimate model parameters generally leads to smaller HC standard errors compared to ordinary least squares (OLS), which translates into improved inference in the form of shorter confidence intervals and more powerful hypothesis tests. In an extensive empirical analysis based on historical stock returns and commonly used factors, we find that conditional heteroskedasticity is pronounced and that WLS and ALS can dramatically shorten confidence intervals compared to OLS, especially during times of financial turmoil.

1. Introduction

The Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) is a cornerstone of finance and marks the birth of asset pricing theory. It is part of just about any finance curriculum in academia and is also widely used in the industry; for an in-depth review, see Fama and French (2004). The model states that the expected excess return of a stock is proportional to the expected excess return of the market:

$$\mathbb{E}(R - r_f) = \beta \mathbb{E}(R_m - r_f) ,$$

where R denotes the return of the stock, r_f denotes the risk-free rate, and R_m denotes the return of the market. The coefficient β is known as the "beta" of the stock and measures its riskiness with respect to the market: the larger is β, the riskier is the stock compared to the market and the larger is also the expected (excess) return of the stock.

In practice, the beta of a stock is unknown and needs to be estimated from historical data. To this end it is common to consider a regression model of the following kind:

$$r_t = \alpha + \beta r_{m,t} + \epsilon_t \ , \tag{1.1}$$

with $\mathbb{E}(\epsilon_t | r_{m,t}) = 0$. Here, $t \in \{1, \dots, n\}$ indexes dates,

- $r_{f,t}$ denotes the return of the risk-free asset at date t,

* Corresponding author at: Department of Economics, University of Zurich, 8032 Zurich, Switzerland.
E-mail addresses: elliot.beck@bf.uzh.ch (E. Beck), gianluca.denard@econ.uzh.ch (G. De Nard), michael.wolf@econ.uzh.ch (M. Wolf).

https://doi.org/10.1016/j.iref.2023.03.009
Received 22 March 2022; Received in revised form 1 March 2023; Accepted 6 March 2023
Available online 25 March 2023
1059-0560/

- $r_t := R_t - r_{f,t}$ denotes the excess return of the stock at date t,
- $r_{m,t} := R_{m,t} - r_{f,t}$ denotes the excess return of the market at date t, and
- ε_t denotes a mean-zero error term.

Even though the CAPM postulates that $\alpha = 0$ in regression (1.1), it still is customary to include an intercept in practice when estimating the model. One of the reasons is that one might be interested in testing a violation of the CAPM, that is, in testing the hypothesis $H_0 : \alpha = 0$ against the alternative $H_1 : \alpha \neq 0$. Another reason is that the usual interpretation of the R^2 statistic of an OLS regression (as the percentage of the variation in the regressand explained by the estimated model) is not valid if the regression does not contain an intercept.

More generally, the Arbitrage Pricing Theory (APT) of Ross (1976) states that the expected excess return of a stock can be modeled as a linear function of *several* factors or theoretical market indices. Thereby, the sensitivity to changes in each factor is represented by a factor-specific beta coefficient. In slight abuse of notation, a general (multi-)factor model can be written as

$$r_t = \beta' x_t + \varepsilon_t \ , \tag{1.2}$$

where

- $x_{k,t}$ denotes the return of factor k at date t, stacked into $x_t := (x_{1,t}, \dots, x_{K,t})'$,
- $x_{1,t} \equiv 1$ in case an intercept is included,
- β_k denotes the beta of factor k, stacked into $\beta := (\beta_1, \dots, \beta_K)'$, and
- ε_t denotes a mean-zero error term.

Clearly, model (1.2) nests model (1.1) with the choices $x_t := (1, r_{m,t})'$ and $\beta := (\alpha, \beta)'$. For reasons that will become apparent below, it is more convenient for our purposes to include a (potential) intercept in x_t, in which case its coefficient is β_1, as opposed to 'listing' it separately, with coefficient α; of course, in such a case one would not think of the intercept as an actual factor. For the same reasons, it is more convenient to denote the vector of factors by x_t rather than by f_t, with the latter convention being more standard in the literature.

The search for factors that explain the cross-section of expected stock returns has produced hundreds of potential candidates. Both Green et al. (2013) and Harvey et al. (2016) find more than 300 articles and factors in this strand of literature. Additionally, Cochrane (2011) and more recently (McLean & Pontiff, 2016) state that we have a "zoo" of (new) factors. Note that Hou et al. (2017) even replicate the entire anomalies literature in finance and accounting by compiling a largest-to-date data library that contains 447 anomaly variables.

Arguably, the classic multi-factor model is the three-factor model of Fama and French (1993):

$$x_t := (1, r_{m,t}, \mathrm{SMB}_t, \mathrm{HML}_t)' \ , \tag{1.3}$$

where SMB denotes the size factor and HML the value factor. The Fama–French three-factor model was extended to a four-factor model by Carhart (1997):

$$x_t := (1, r_{m,t}, \mathrm{SMB}_t, \mathrm{HML}_t, \mathrm{UMD}_t)' \ , \tag{1.4}$$

where UMD denotes the momentum (winners minus losers) factor, and recently to a five-factor model by Fama and French (2015):

$$x_t := (1, r_{m,t}, \mathrm{SMB}_t, \mathrm{HML}_t, \mathrm{RMW}_t, \mathrm{CMA}_t)' \ , \tag{1.5}$$

where RMW denotes the profitability factor, and CMA the investment factor. There are of course many other (multi-) factor models, but to make our point clear, and as it is not obvious from the literature which and how many factors should be considered, we will focus on the most common ones listed above.

The parameter vector β in a factor model is typically estimated via ordinary least squares (OLS). To this end, it is standard to use daily data with the most common samples sizes being $252 \leq n \leq 1260$, that is, one to five years of past data; for example, see Frazzini and Pedersen (2014) or De Nard et al. (2021). Alternatively, Bloomberg uses two years of weekly data for their beta estimates and some even use monthly data with the most common samples sizes being $n = 60$ or $n = 120$, that is, five or ten years of past data; for example, see Damodaran (2012, Chapter 8) and Stock and Watson (2019, Section 4.2).

Some researchers still assume that stock and factor returns are independent and identically distributed (i.i.d.) through time; for example, see Campbell et al. (2012, Section 4.3). It is more general, and more realistic, however, to assume that stock and factor returns are (strictly) stationary through time. Even this weaker assumption implies that the error terms are *unconditionally homoskedastic*, that is, $\mathbb{E}(\varepsilon_t^2)$ is a constant number and does not depend on t. A common occurrence, which tends to be ignored by many applied researchers, is the one of *conditional heteroskedasticity* of the error terms. In our general formulation (1.2), which will be the basis of our analysis from here on (unless otherwise stated), this means that $\mathbb{E}(\varepsilon_t^2 | x_t)$ in general is not a constant number but a function of x_t.

2. Dealing with conditional heteroskedasticity

2.1. Methodology

Conditional heteroskedasticity does not present a problem for the estimation of model (1.2) via OLS in the sense that the OLS estimator of β is still consistent under weak regularity conditions; having said this, the OLS estimator is no longer efficient in the sense of having the smallest asymptotic covariance matrix.

On the other hand, conditional heteroskedasticity does present a problem for inference in model (1.2) in the sense that the usual standard errors of the OLS estimators of linear combinations of β (such as specific elements of β) are no longer valid, since these standard errors are based on an assumption of conditional homoskedasticity. Here, by the "usual" standard errors we mean the default textbook standard errors; for example, see Hayashi (2000, Section 2.6).

The common way to deal with this problem is to combine OLS estimation with heteroskedasticity consistent (HC) standard errors, which guarantees asymptotically valid inference under conditional heteroskedasticity of unknown form. Such HC standard errors go back to the seminal paper of White (1980) but, importantly, there have been subsequent alternative proposals to deliver better finite-sample performance; for example, see Romano and Wolf (2017, Section 4) who describe in detail five versions of HC standard errors (HC0–HC4) and recommend HC3 standard errors for practical use.

As an alternative to OLS, Romano and Wolf (2017) suggest to use weighted least squares (WLS) or adaptive least squares (ALS). These methods are based on the concept of a *skedastic function* that maps the factor (vector) x_t into the corresponding conditional variance of the error term:

$$v(x_t) := \mathbb{E}(\varepsilon_t^2 | x_t) \ .$$

This function is unknown in practice but can be estimated from the observed data, resulting in an estimator $\hat{v}(\cdot)$; see below for a specific proposal.

The WLS method weights the data by division by $\sqrt{\hat{v}(x_t)}$ before applying OLS. That is, one considers the 'transformed' regression model

$$\frac{r_t}{\sqrt{\hat{v}(x_t)}} = \beta' \frac{x_t}{\sqrt{\hat{v}(x_t)}} + \varepsilon_t^* \quad \text{with} \quad \varepsilon_t^* := \frac{\varepsilon_t}{\sqrt{\hat{v}(x_t)}} \ . \tag{2.1}$$

The parameter vector β is identical in model (2.1) compared to the original model (1.2), otherwise the exercise of transforming the model would be pointless. Importantly, one also needs to use HC standard errors for the inference in model (2.1) to allow for the possibility that $\hat{v}(\cdot)$ may not be a consistent estimator of the true skedastic function, as explained in detail by Romano and Wolf (2017).

The ALS method 'decides' between the OLS method and the WLS method based on a pre-test for conditional homoskedasticity. Only if this test rejects the null, that is, if this tests detects a significant amount of conditional heteroskedasticity in the data, does one use WLS; otherwise one uses OLS. Of course, either way, one must use corresponding HC standard errors for the inference.

In Monte Carlo studies, Romano and Wolf (2017) demonstrate two advantages of WLS and ALS over OLS in the presence of conditional heteroskedasticity. First, the point estimators tend to have smaller mean squared error (MSE); second, the HC3 standard errors for the point estimators tend to be smaller, resulting in shorter confidence intervals and more powerful hypothesis tests. Further empirical evidence is provided in Sterchi and Wolf (2017).

2.2. Parametric specification of the skedastic function

We use the following parametric specification for the skedastic function:

$$v_\theta(x_t) := \exp(\nu + \gamma_2 |x_{2,t}| + \cdots + \gamma_K |x_{K,t}|) \quad \text{with} \quad \theta := (\nu, \gamma_2, \ldots, \gamma_K)' \ . \tag{2.2}$$

This specification tacitly assumes that an intercept is included in the factor model, that is, $x_{1,t} \equiv 1$; otherwise the specification should be

$$v_\theta(x_t) := \exp(\nu + \gamma_1 |x_{1,t}| + \cdots + \gamma_K |x_{K,t}|) \quad \text{with} \quad \theta := (\nu, \gamma_1, \ldots, \gamma_K)' \ . \tag{2.3}$$

Specification (2.2) without the absolute values around the $x_{k,t}$ is proposed by Wooldridge (2012, Chapter 8). Since in our case the $x_{k,t}$ can take on both negative and positive values, and it is reasonable to assume that conditional heteroskedasticity depends on the magnitude only, using absolute values makes more sense.

Another specification proposed by Romano and Wolf (2017) is

$$v_\theta(x_t) = \exp(\nu + \gamma_2 \log|x_{2,t}| + \cdots + \log \gamma_K |x_{K,t}|) \ .$$

But this specification is problematic when not all of the $|x_{k,t}|$ are bounded away from zero, which is clearly the case in our context.

In order to estimate the parameter vector θ in (2.2), we first estimate model (1.2) via OLS and denote the corresponding residuals by $\hat{\varepsilon}_t$. Then, in principle, we would estimate the following regression by OLS:

$$\log\left[\max(\delta^2, \hat{\varepsilon}_t^2)\right] = \nu + \gamma_2 |x_{2,t}| + \cdots + \gamma_K |x_{K,t}| + u_t$$

and denote the resulting estimator by $\hat{\theta} := (\hat{\nu}, \hat{\gamma}_2, \ldots, \hat{\gamma}_K)'$.

The introduction of the lower bound δ^2 on the left-hand side in this regression is necessary in order to avoid taking the log of values very close (or even equal) to zero. Romano and Wolf (2017) recommend the generic choice $\delta = 0.1$ but we find that, for many data sets, truncation at $\delta^2 = 0.01$ takes place in a large fraction of the observations, which is not conducive to an accurate estimation of θ. If necessary, we therefore 'blow up' all the variables in the regression by a factor of ten until the fraction of truncations is at most 5%. More specifically, the parameter vector θ is estimated by the OLS regression

$$\log\left[\max(\delta^2, 10^q \hat{\epsilon}_t^2)\right] = \nu 10^q + \gamma_2 |10^q x_{2,t}| + \cdots + \gamma_K |10^q x_{K,t}| + u_t \ , \tag{2.4}$$

where $q \in \{0, 1, 2, \ldots\}$ is the smallest non-negative integer such that the fraction of truncations on the left-hand side is at most 5% with the lower bound $\delta^2 = 0.01$. With the resulting $\hat{\theta} := (\hat{\nu}, \hat{\gamma}_2, \ldots, \hat{\gamma}_K)'$ in hand, the estimator of the skedastic function used in regression (2.1) is then given by $\hat{v}(\cdot) := v_{\hat{\theta}}(\cdot)$.

Regression (2.4) also determines the 'decision' underlying for the ALS estimator and the corresponding inference: If the (joint) null hypothesis $H_0 : \gamma_2 = \cdots = \gamma_K = 0$ is rejected in this regression at significance level 0.1, then ALS coincides with WLS; otherwise, it coincides with OLS.

Remark 2.1. Free programming code in the R language that implements this methodology can be downloaded at www.econ.uzh.ch/en/people/faculty/wolf/publications.html.

3. Theoretical analysis

Romano and Wolf (2017) assume i.i.d. data to prove (asymptotic) validity of inference based on WLS or ALS. But making such an assumption is unrealistic in the context of financial returns, for example because of the well-known phenomenon of *volatility clustering*, at least at shorter horizons such as at the daily or at the weekly horizon. For this reason we need to extend the methodology of Romano and Wolf (2017) by proving its validity under a more general set set of assumptions that is realistic for financial returns (at least when the assets are stocks and commonly used factors).

We maintain the following set of assumptions throughout the paper:

(A1) The linear model is of the form

$$r_t = x_t'\beta + \epsilon_t \quad (t = 1, \ldots, n) \ , \tag{3.1}$$

where $x_t \in \mathbb{R}^K$ is a vector of explanatory variables (regressors) possibly including a constant, $\beta \in \mathbb{R}^K$ is a coefficient vector, and ϵ_t is the unobservable error term with certain properties to be specified below.

(A2) The sample $\left\{(r_t, x_t')\right\}_{t=1}^n$ is strictly stationary and ergodic.

(A3) The error terms satisfy

$$\mathbb{E}(\epsilon_t | x_t, \ldots, x_1, \epsilon_{t-1}, \ldots, \epsilon_1) = 0 \quad \forall t \ . \tag{3.2}$$

(A4) The $K \times K$ matrix $\Sigma_{xx} := \mathbb{E}(x_t x_t')$ is nonsingular (and hence finite). Furthermore, $\sum_{t=1}^n x_t x_t'$ is invertible with probability one.

(A5) The $K \times K$ matrix $\Omega := \mathbb{E}(\epsilon_t^2 x_t x_t')$ is nonsingular (and hence finite).

(A6) There exists a nonrandom function $v : \mathbb{R}^K \to \mathbb{R}_+$ such that

$$\mathbb{E}(\epsilon_t^2 | x_t) = v(x_t) \ . \tag{3.3}$$

Therefore, the *skedastic function* $v(\cdot)$ determines the functional form of the conditional heteroskedasticity. Note that under (A6),

$$\Omega = \mathbb{E}\left[v(x_t) \cdot x_t x_t'\right] \ .$$

The two generalizations compared to Romano and Wolf (2017) are Assumptions (A2)–(A3), the remaining assumptions being identical. This new set of assumptions allows for time-series dynamics that are realistic, or at least plausible, for many financial returns; in particular, time-varying conditional (co)-volatilities can be incorporated, such as (multivariate) GARCH dynamics.

It is useful to introduce the customary vector-matrix notations

$$r := \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}, \quad \epsilon := \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}, \quad X := \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} = \begin{bmatrix} x_{11} & \ldots & x_{1K} \\ \vdots & \ldots & \vdots \\ x_{n1} & \ldots & x_{nK} \end{bmatrix},$$

so that Eq. (3.1) can be written more compactly as

$$r = X\beta + \epsilon \ . \tag{3.4}$$

Furthermore, Assumptions (A2), (A3), and (A5) imply that

$$\text{Var}(\epsilon | X) = \begin{bmatrix} v(x_1) & & \\ & \ddots & \\ & & v(x_n) \end{bmatrix} .$$

We now can state the following theorem.

Theorem 3.1. *With the above re-definitions of Assumptions (A.2)–(A.3), the following results of Romano and Wolf (2017) continue to hold:*

- *Lemma 3.1*
- *Corollary 3.1*
- *Theorem 3.1*
- *Theorem 4.1*
- *An analog of Verification B.2 of assumptions for the parametric specification $v_\theta(\cdot)$ of (2.2)*

Wherever necessary, the role of y_i in these results is now taken over by r_t, which is just a different notation for the response variable in the regression model (3.1). Also, to point out the obvious, observations are now indexed by t instead of by i as in Romano and Wolf (2017). The proof of the theorem is deferred to Appendix.

For the detailed statements of the five results listed in Theorem 3.1 the reader is referred to Romano and Wolf (2017). But to make this paper (more) self-contained in terms of grasping the essentials for practical applications, we now briefly describe the import and the implications of the various results.

Given two real-valued functions $a(\cdot)$ and $b(\cdot)$ defined on $\mathbb{R}^K$ (the space where x_i lives), define $\Omega_{a/b}$ to be the matrix given by

$$\Omega_{a/b} := \mathbb{E}\left[\frac{a(x_i)}{b(x_i)} \cdot x_i x_i'\right] ,$$

assuming the expectation exists, of course. By the final (that is, fifth) result listed in Theorem 3.1, $\hat{\theta}$ converges in probability to a limiting non-stochastic value θ_0 for the parametric specification $v_\theta(\cdot)$ of (2.2); recall here that $\hat{\theta}$ is the OLS estimator of θ based on the linear model (2.4). For compactness of notation let $w := v_{\theta_0}$, and recall that v denotes the true skedastic function.

Under the stated assumptions (A.1)–(A.6) and some further moment assumptions, it then follows from the first three results listed in Theorem 3.1 that the WLS estimator is asymptotically normal:

$$\sqrt{n}(\hat{\beta}_{\mathrm{WLS}} - \beta) \xrightarrow{d} N\left(0, \mathrm{Avar}(\hat{\beta}_{\mathrm{WLS}})\right) \quad \text{with} \quad \mathrm{Avar}(\hat{\beta}_{\mathrm{WLS}}) := \Omega_{1/w}^{-1}\Omega_{v/w^2}\Omega_{1/w}^{-1} ,$$

where $\xrightarrow{d}$ denotes convergence in distribution. Of course, this result implies that the WLS estimator is consistent, that is,

$$\hat{\beta}_{\mathrm{WLS}} \xrightarrow{p} \beta ,$$

where $\xrightarrow{p}$ denotes convergence in probability.

Furthermore, under some moment conditions, it follows from the fourth result listed in Theorem 3.1 that the asymptotic covariance matrix $\mathrm{Avar}(\hat{\beta}_{\mathrm{WLS}})$ can be estimated consistently by applying standard HC technology to the weighted data specified in (2.1); for example, see Long and Ervin (2000). Inference on β can therefore be based on $\hat{\beta}_{\mathrm{WLS}}$ in conjunction with a consistent HC estimator $\widehat{\mathrm{Avar}}(\hat{\beta}_{\mathrm{WLS}})$ of $\mathrm{Avar}(\hat{\beta}_{\mathrm{WLS}})$ by applying the usual 'textbook formulas' for the t-test, for the F-test, and for confidence intervals. The fourth result listed in Theorem 3.1, together with the discussion just below it, implies that such inference is asymptotically valid; for example, a correct null hypothesis will be rejected with a probability that is bounded above by the nominal significance level in the limit[1]; and a confidence interval will contain the true parameter with a probability that converges to the nominal coverage level in the limit.

Analogously, consistency of the ALS estimator $\hat{\beta}_{\mathrm{ALS}}$ and asymptotic validity of the inference on β based on $\hat{\beta}_{\mathrm{ALS}}$ is established as well.

4. Empirical analysis

4.1. Data and model construction

We download daily stock return data from the Center for Research in Security Prices starting on January 1, 1964, and ending on December 31, 2019. We restrict attention to stocks from the NYSE, AMEX, and NASDAQ stock exchanges. We also download the daily risk-free rate and the returns on the five factors of Fama and French (2015) and the momentum (winners minus losers) factor of Carhart (1997) during the same period from the website of Ken French.

We restrict our attention to a one-factor and three multi-factor models: the CAPM, which uses $x_t := (1, r_{m,t})'$; the three- and five-factor model of Fama and French (1993, 2015), which uses x_t as defined in (1.3), respectively in (1.5); and the four-factor model of Carhart (1997) which uses x_t as defined in (1.4).[2] The models are re-estimated once a year on December 31, starting in 1968 and ending in 2019. Doing so results in a total of 52 yearly estimates, indexed by $h = 1, \ldots, 52$. For any h, the models are estimated

[1] When the parameter under test is univariate and the test is one-sided, the limiting rejection probability typically is below the nominal significance level, unless the parameter is on the boundary of the null space.

[2] To save space and avoid repetitiveness, we sometimes only show results for the largest factor model, hence the five-factor model of Fama and French (2015), as the results for the other two factor models are very similar.

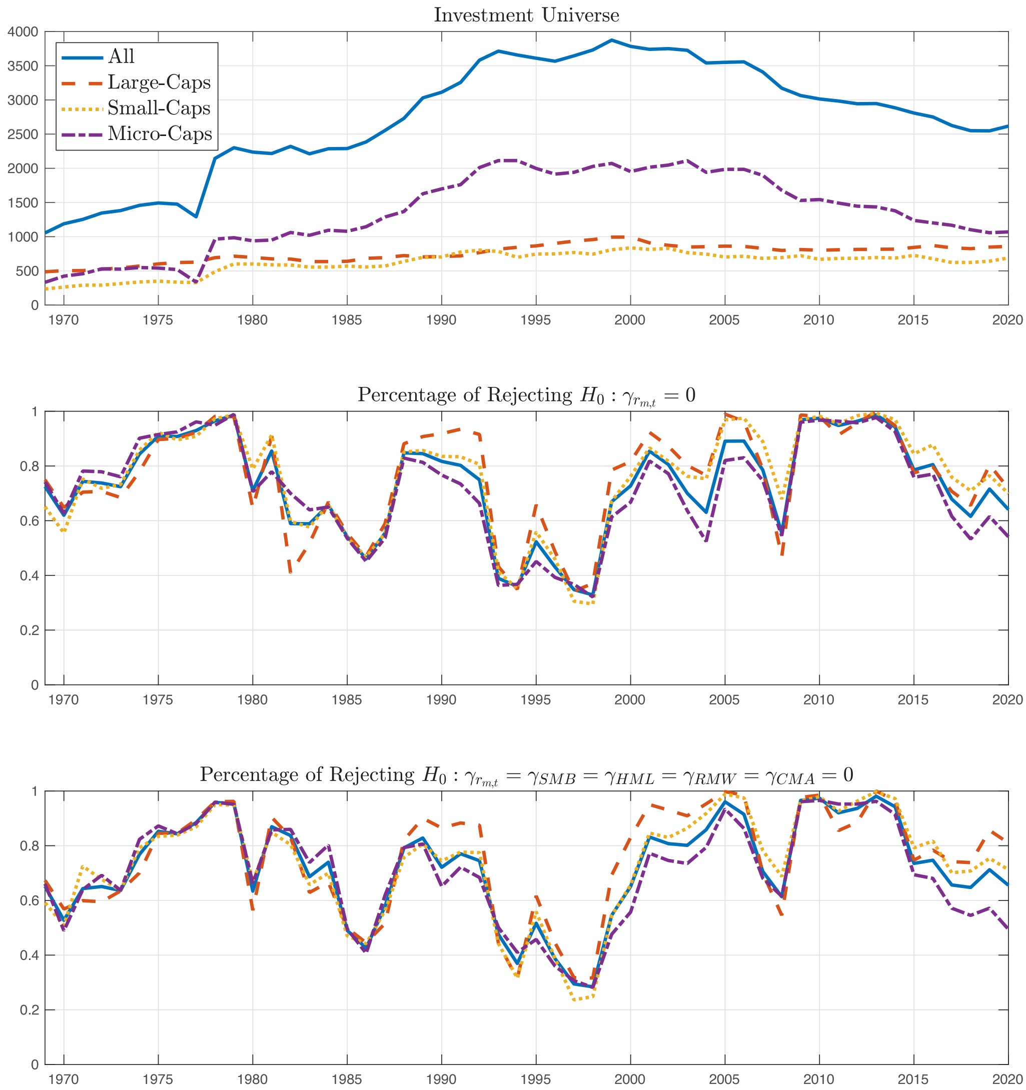

Fig. 1. The panel on the top plots the yearly time series of the (size of the) investment universe, and its decomposition into large-, small-, and micro-cap stocks. The panel in the middle plots the percentages of rejecting the null of conditional homoskedasticity in the CAPM for the entire investment universe and its decomposition. The panel on the bottom plots the percentages of rejecting the null of conditional homoskedasticity in the Fama–French five-factor model for the entire investment universe and for its decomposition.

based on the most recent $n = 1260$ daily returns, which roughly corresponds to using five years of past data; therefore, we are using a rolling-window rather than an expanding-window approach, which is standard in the literature. For any h, the investment universe is comprised of the set of stocks that have a complete return history over the corresponding past 1260 days. The top panel of Fig. 1 displays the resulting size of the investment universe over time (N_h), together with its decomposition into large-, small-, and micro-cap stocks.[3] The middle panel of the figure displays the percentage of stocks for which significant conditional heteroskedasticity is detected in the CAPM, for the entire universe and also for the three sub-universes; the bottom panel does the same for the Fama–French five-factor model. The resulting message is loud and clear: There is ample evidence for conditional heteroskedasticity; indeed, the percentages are generally well above 0.5 and can even get very close to 1, such as in the years after the financial crisis of 2008.

[3] In principle, we could allow for small percentage of missing returns during the estimation period, to be replaced by zeros, which would result in even larger investment universe. But even with our 'strict' rule the investment universes are large, as the figure shows.

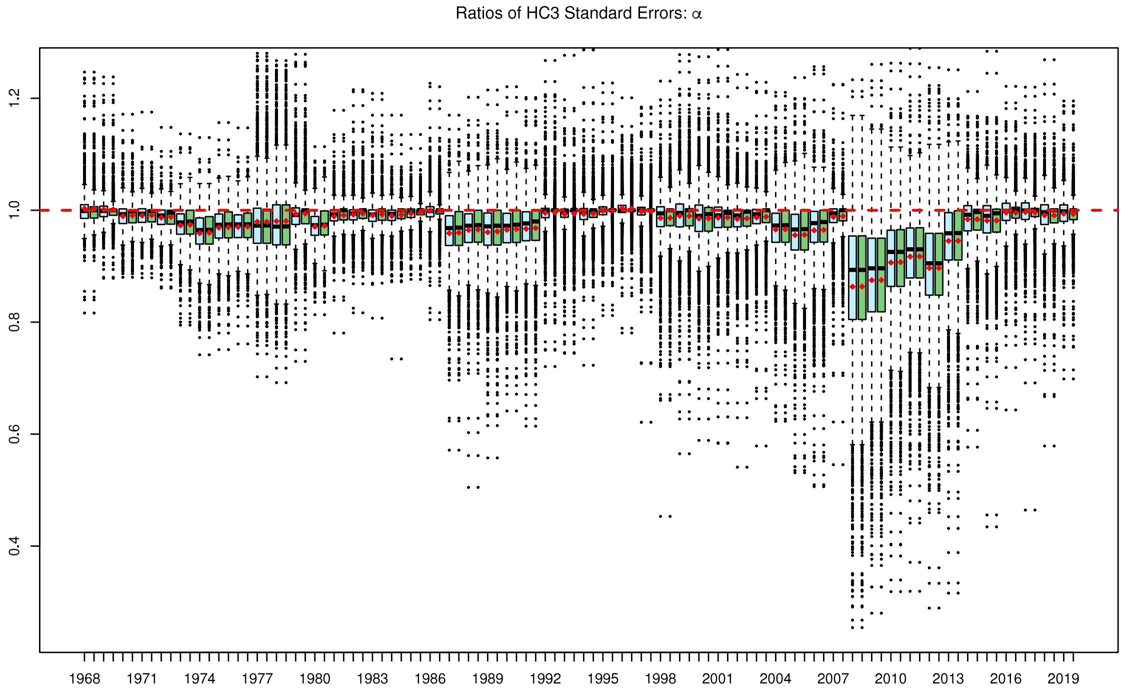

Fig. 2. CAPM boxplots for the two ratios defined in (4.1) over 52 years for the parameter of interest $\alpha := \beta_1$. For any year, we plot first the WLS/OLS boxplots in **blue**, followed by the ALS/OLS boxplots in **green**. In each box, the bar indicates the median ratio whereas the diamond-shaped symbol indicates the average ratio. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Performance measure

Since WLS is based on an OLS regression after weighting the data, the point estimates for β differ between WLS and OLS. In their Monte Carlo study, Romano and Wolf (2017) show that WLS and ALS estimators for specific entries of β typically have smaller mean squared error (MSE) than the OLS estimator; further numerical evidence is provided by Sterchi and Wolf (2017). Unfortunately, in an application to real data, the MSE values cannot be compared, since the true parameters are unknown.

Therefore, we restrict the comparison to the resulting standard errors when inference for univariate parameters is carried out. In the following, we restrict attention to an arbitrary element β_k of β; more generally, inference for (non)linear combinations of β could be considered as well. Note that the ratio of any two standard errors is equal to the ratio of the lengths of the two corresponding confidence intervals, as the confidence intervals that we consider are the usual 'textbook' ones. Specifically, a generic nominal $1-\lambda$ confidence interval for β_k is given by

$$\hat{\beta}_{k,*\mathrm{LS}} \pm t_{n-K,1-\frac{\lambda}{2}} \times \mathrm{SE}_{\mathrm{HC}}(\hat{\beta}_{k,*\mathrm{LS}}) \ ,$$

where $*\mathrm{LS} \in \{\mathrm{OLS}, \mathrm{WLS}, \mathrm{ALS}\}$, $t_{n-K,\lambda}$ denotes the λ quantile of the t_{n-K} distribution, and $\mathrm{SE}_{\mathrm{HC}}$ denotes a HC3 standard error. Consequently, the length of a confidence interval is proportional to the underlying standard error irrespective of the nominal level $1-\lambda$.

For any year we compute the two ratios

$$\frac{\mathrm{SE}_{\mathrm{HC}}(\hat{\beta}_{k,\ \mathrm{WLS}})}{\mathrm{SE}_{\mathrm{HC}}(\hat{\beta}_{k,\ \mathrm{OLS}})} \quad \text{and} \quad \frac{\mathrm{SE}_{\mathrm{HC}}(\hat{\beta}_{k,\ \mathrm{ALS}})}{\mathrm{SE}_{\mathrm{HC}}(\hat{\beta}_{k,\ \mathrm{OLS}})} \ , \tag{4.1}$$

and then use boxplots to visually 'summarize' their distribution over the 52 years.[4] A ratio larger (smaller) than one implies that the HC3 standard error for WLS, respectively ALS, is larger (smaller) than the corresponding standard error for OLS. For both WLS and ALS, we report the percentage of ratios that are smaller than one, and thus the percentage of cases where WLS, respectively ALS, leads to improved inference compared to OLS.

Remark 4.1. Comparing the (average) lengths of confidence intervals computed from two different inference methods could, in principle, be misleading, namely if one inference method is valid, in the sense of producing confidence intervals whose true coverage

[4] Note that each boxplot is based on the time-varying investment universe depicted in Fig. 1.

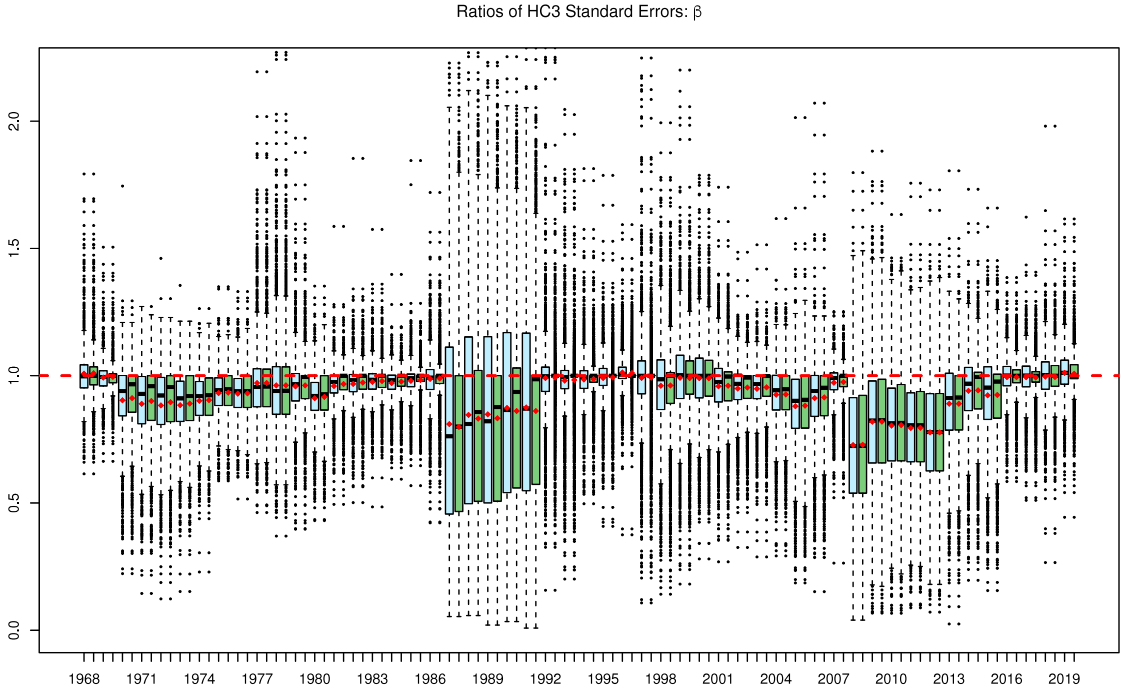

Fig. 3. CAPM boxplots for the two ratios defined in (4.1) over 52 years for the parameter of interest $\beta := \beta_2$. For any year, we plot first the WLS/OLS boxplots in **blue**, followed by the ALS/OLS boxplots in **green**. In each box, the bar indicates the median ratio whereas the diamond-shaped symbol indicates the average ratio. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

probability is equal to (or larger) than the nominal confidence level, whereas the other one is not. But this is not a concern here, since we are estimating linear regression models with either $K = 2$, $K = 4$, $K = 5$ or $K = 6$ regressors based on samples of size of $n = 1260$, in which case finite-sample validity is, for all practical purposes, guaranteed for inference based on OLS, WLS, and ALS. ■

We also report the results on the pretest for conditional homoskedasticity, namely the percentage of times the null hypothesis was rejected. These results shed light on whether (statistically significant) conditional heteroskedasticity is indeed a common occurrence in financial factor-model regressions.

4.3. Results: CAPM

First, we focus on $\alpha := \beta_1$ and look at the entire investment universe. Fig. 2 shows that for both methods, WLS and ALS, the median ratio (4.1) is *always* (weakly) below one. Furthermore, in most of the 52 years roughly 75% of the ratios (or more) lie weakly below one. Additionally, we find that for 49 out of the 52 years, the mean ratio lies below one. This is a remarkable finding because, by construction, (potential) small outliers of the ratios are bounded below by zero, but (potential) large outliers are unbounded, so that outliers should move the mean in the direction "above one". The bottom line is that using WLS or ALS generally results in shorter confidence intervals for α compared to using OLS. This is particularly the case during times of financial turmoil (e.g., Black Monday in 1987 or the financial crisis in 2008); intuitively, this finding can be explained by the fact that after a stock market crash or during a financial crisis, volatility clustering is a (more) pronounced phenomenon, especially at the daily return frequency as considered here.

Second, we focus on $\beta := \beta_2$ and look again at the entire investment universe. The results are similar to the ones for α. Fig. 3 shows that (i) for both methods, WLS and ALS, the median ratio (4.1) is also always (weakly) below one; (ii) in most of the 52 years at least 75% of the ratios lie (weakly) below one; and (iii) even most of the mean ratios lie (weakly) below one. It is worth to mention that for most of the periods the results for ALS are a bit more 'condensed' compared to those for WLS. In particular in the 80s and 90s, the 75th percentile of ALS avoids ratios in excess of one. Therefore, using WLS and especially ALS generally results in shorter confidence intervals for β compared to using OLS. As already discussed above, the benefit of WLS and ALS increases during times of financial turmoil due to the higher degree of conditional heteroskedasticity then.

As a robustness check, motivated by Fama and French (2008), we now repeat the analysis by breaking up the entire investment universe into large-, small-, and micro-cap stocks. For a particular year, a stock is classified as 'large' if it has a market cap above the 50th NYSE percentile; as 'small' if it has a market cap between the 20th and 50th NYSE percentile; and as 'micro' if it has a market cap below the 20th NYSE percentile. The resulting 'sub-universe sizes' are displayed in the top panel of Fig. 1. We point out

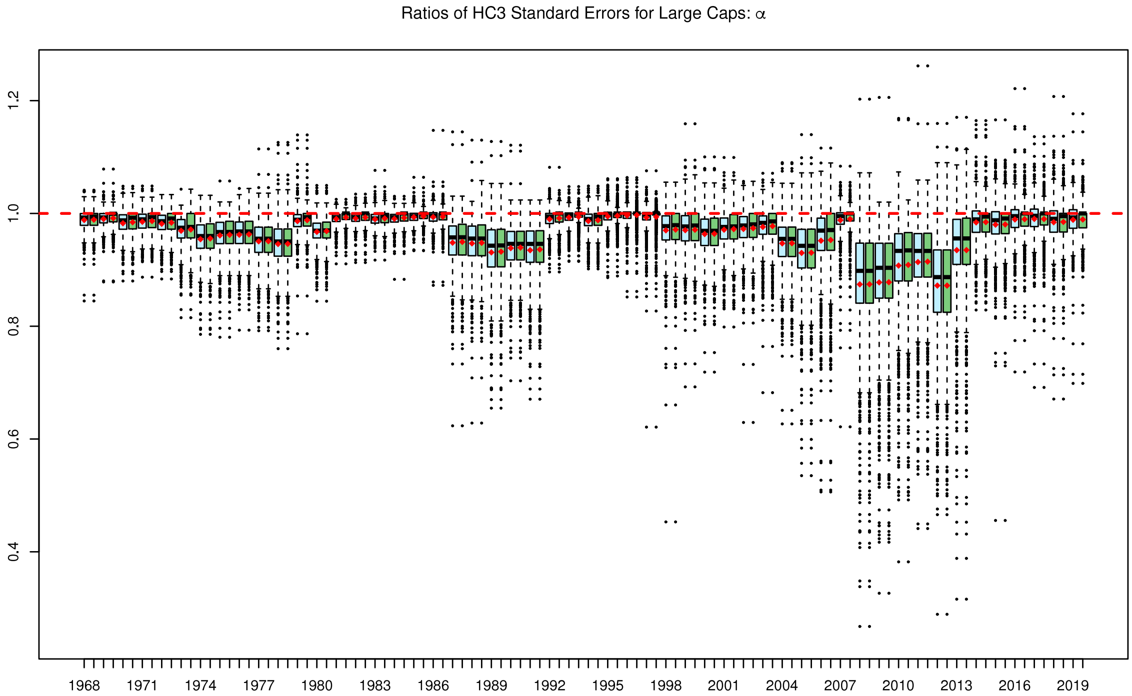

Fig. 4. Similar to Fig. 2 except that we now restrict attention to large-cap stocks.

that although micro-cap stocks generally have the largest sub-universe size (on average as large as the other two sizes together), they only make up a minor fraction of the total market capitalization (on average about 3%).

Figs. 4 and 5 are the equivalents of Figs. 2 and 3 when attention is restricted to large-cap stocks. One can see that, in a given year, the ratios now show less dispersion: In general the length of the boxes is reduced and the number of outliers as well. The benefit of using WLS/ALS during times of financial turmoil is even more pronounced now; for example, the mean ratios for both WLS and ALS are down to about 0.5 in the years after Black Monday 1987 when the parameter of interest is β. We also report the combined results for small- and micro-caps in Fig. 6, where for readability we restrict attention to the ratios for ALS. The benefits compared to OLS are still there, if less pronounced compared to large-cap stocks.

The graph in the middle panel of Fig. 1 sheds some light on the amount of conditional heteroskedasticity present in the CAPM: On average, over the years, the null of conditional homoskedasticity gets rejected roughly 74% of the time.

Finally, is there any preference between WLS and ALS? To address this question, Fig. 7 presents, over the 52 years, the percentage of the standard-error ratios, WLS/OLS and ALS/OLS, that are below one; this done for the entire universe and also for the three sub-universes. According to this metric, ALS uniformly dominates OLS, whereas WLS does not. Therefore, our recommendation in the end is to use ALS.

4.4. Results: Multi-factor models

In this section we analyze the effect of conditional heteroskedasticity in multi-factor models. More specifically, we extend the CAPM with the small minus big (SMB) size factor, the high minus low (HML) value factor, the robust minus weak (RMW) profitability factor, the conservative minus aggressive (CMA) investment factor, and the winners minus loser (UMD) momentum factor. Of course many other factors could be included in the model; however, we restrict our attention to the arguably most common multi-factor models of Fama and French (1993, 2015) and Carhart (1997). In general, we find that also for the multi-factor models conditional heteroskedasticity of the error terms is a common occurrence and that WLS, respectively ALS, tend to reduce HC3 standard errors compared to OLS.

The graph in the bottom panel of Fig. 1 sheds some light on the amount of conditional heteroskedasticity present in the five-factor model: On average, over the years, the null of conditional homoskedasticity gets rejected roughly 75% of the time.

As for the CAPM, we find that usually the null gets rejected more frequently during periods of financial turmoil, as volatility clustering is then more pronounced, and that the percentage of rejecting the null is robust across the investment universes. In terms of the investment universe, we find again that WLS and ALS work across all NYSE breakpoints categories. Thus, for sake of simplicity we report only the results for large-cap stocks.

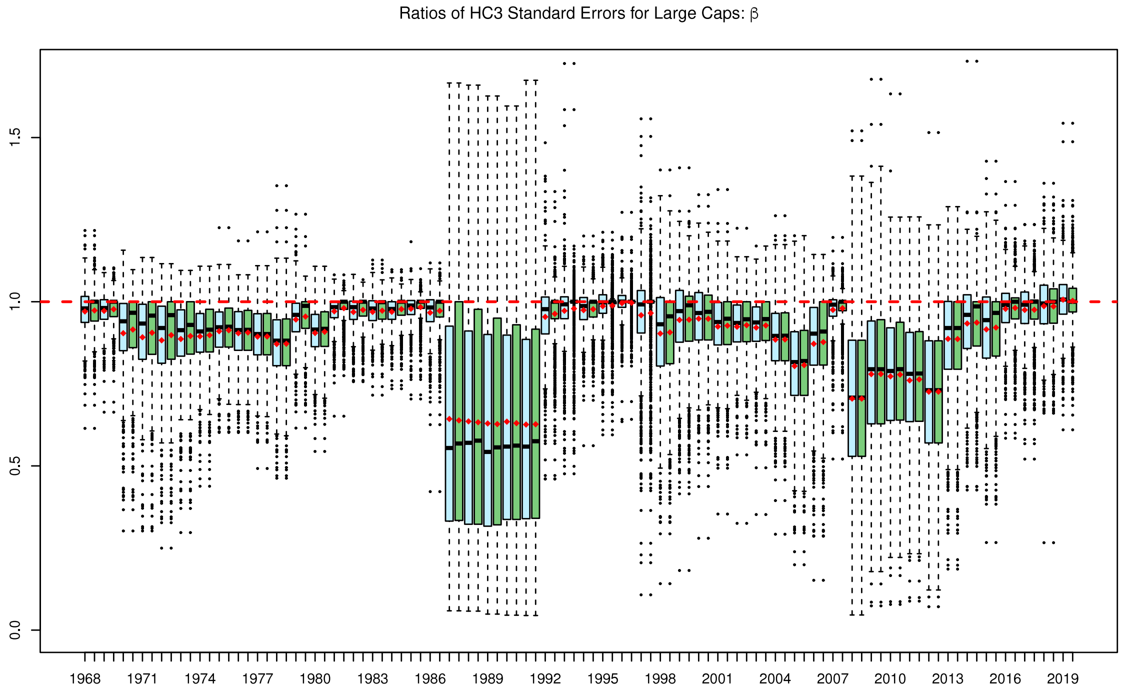

Fig. 5. Similar to Fig. 3 except that we now restrict attention to large-cap stocks.

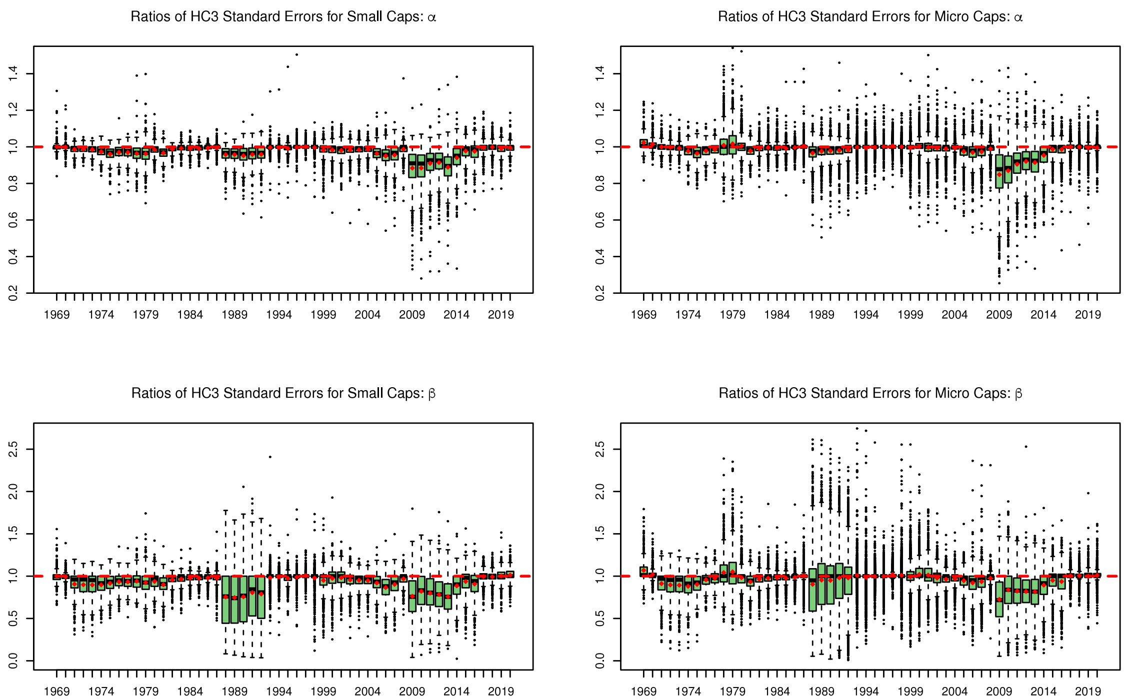

Fig. 6. Similar to Figs. 2 and 3 except that we now restriction attention to small-cap stocks (column on the left) respectively micro-cap stocks (column on the right). To improve readability, the figure only presents boxplots for the ALS/OLS ratios; in unreported results we find similar patterns for the WLS/OLS ratios.

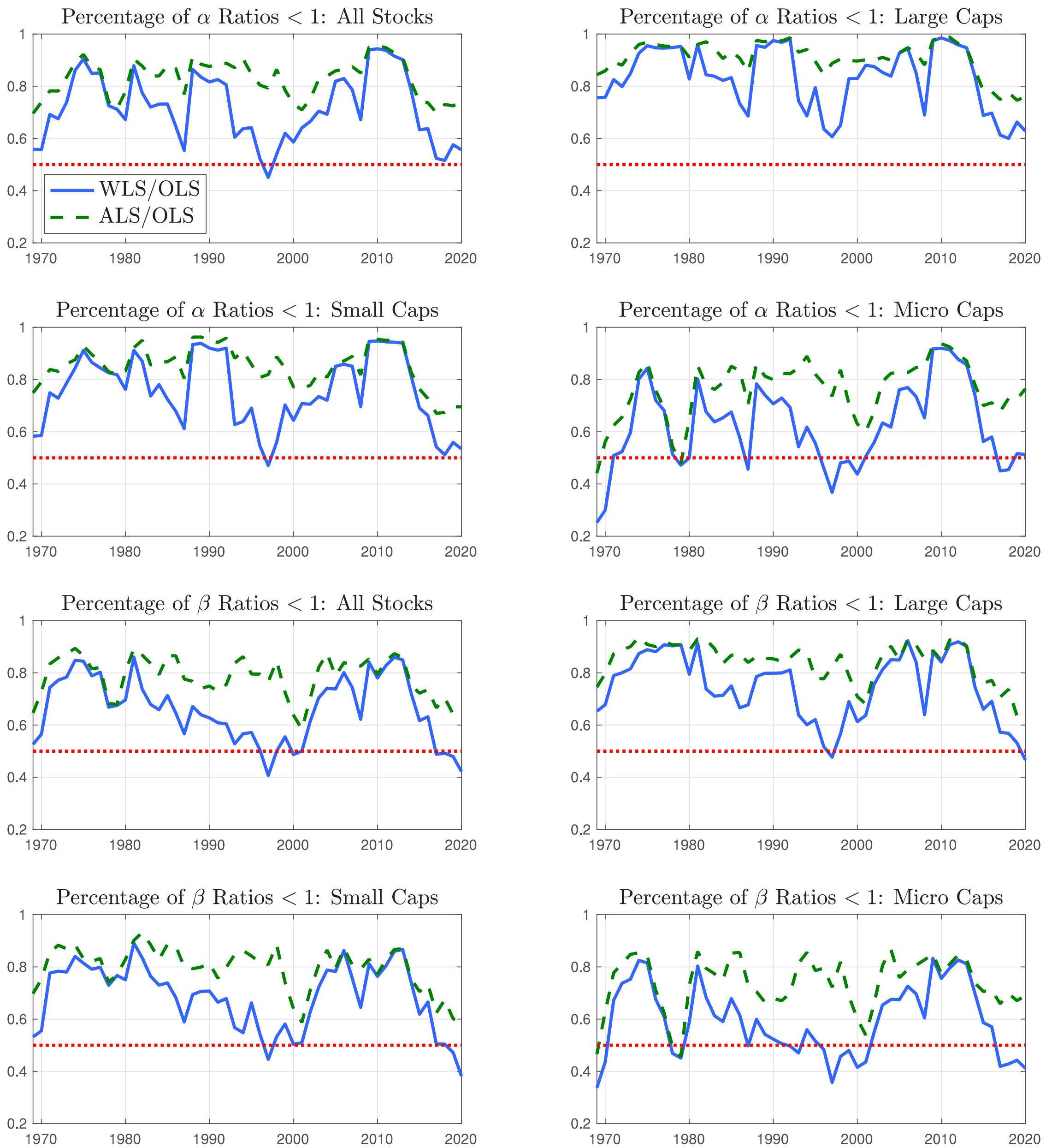

Fig. 7. Percentages of standard-error ratios (4.1) that lie below one for the CAPM. The results are based alternatively on the entire investment universe and on its decomposition into large-, small-, and micro-cap stocks.

Fig. 8 presents the α HC3 standard-error ratios for large-cap stocks. Note that the median and mean α ratios are always below one and that the power of WLS and ALS is slightly higher in the five-factor model compared to the CAPM. We also plot the $\beta_{r_{m,t}}$ HC3 standard-error ratios for large-cap stocks in Fig. 9. The results for the market factor look similar to those for the CAPM, but again they are even slightly better. Finally, we also report the results for the other four Fama–French factors in Fig. 10. In sum, WLS and ALS overall prominently reduce the HC3 standard errors compared to OLS for all investigated factors (and also for the intercept).

Even more impressive are the results concerning the percentages of HC3 standard-error ratios that lie below one, presented in Fig. 11 and Table 1. We find that ALS consistently and often markedly outperforms WLS across all factors, and also for the intercept. Therefore, again, applied researchers are advised to abandon OLS and update instead to ALS, which reduces the HC3 standard errors, and thus the length of the confidence intervals, in 79%–94% of the cases (depending on the coefficient).

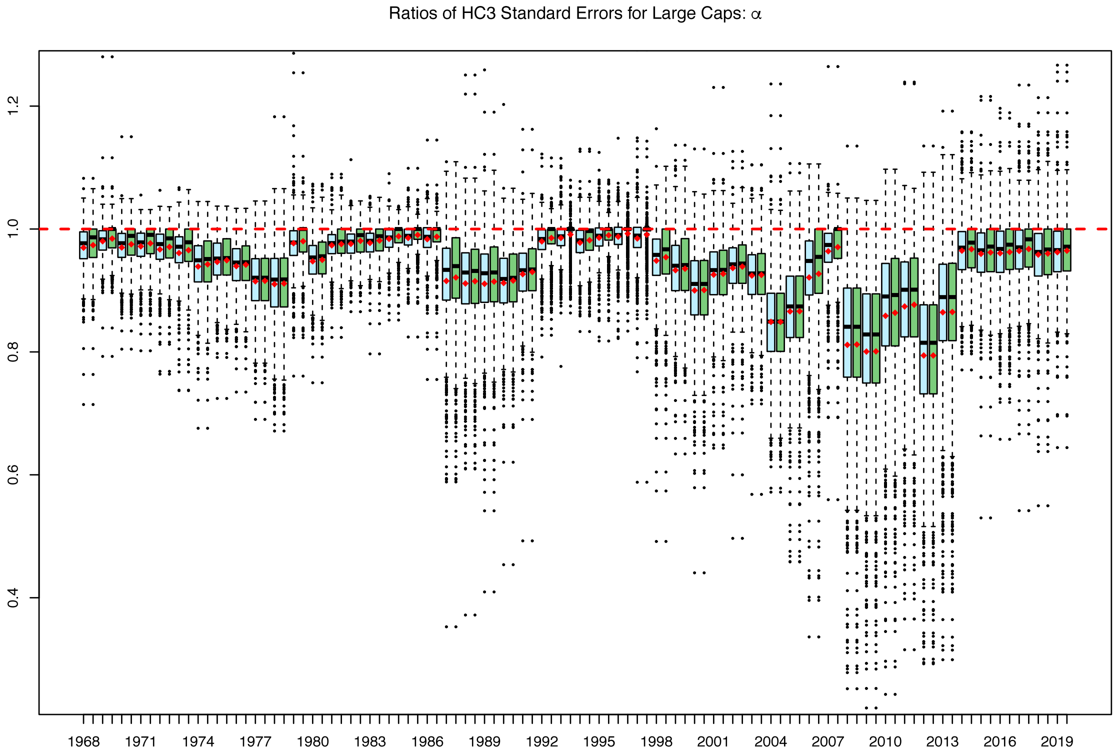

Fig. 8. Five-factor-model boxplots for the two ratios defined in (4.1) for large-cap stocks. For any year, we plot first the WLS/OLS boxplots in **blue**, followed by the ALS/OLS boxplots in **green**. In each box, the bar indicates the median ratio whereas the diamond-shaped symbol indicates the average ratio. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Percentages of HC3 standard-error ratios that lie below one for all coefficients of the various factor models. All percentages are based on daily data for large-cap stocks from 12/31/1968 through 12/31/2019. For any coefficient and model, the highest number appears in **bold face**.

HC3 standard-error ratios < 1							
	α	$\beta_{r_{m,t}}$	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}	β_{UMD}
	CAPM						
WLS	83%	74%	–	–	–	–	–
ALS	**90%**	**83%**	–	–	–	–	–
	Fama–French 3-Factor Model						
WLS	87%	74%	73%	75%	–	–	–
ALS	**93%**	**83%**	**83%**	**84%**	–	–	–
	Carhart 4-Factor Model						
WLS	87%	73%	75%	76%	–	–	69%
ALS	**93%**	**82%**	**83%**	**84%**	–	–	**79%**
	Fama–French 5-Factor Model						
WLS	87%	75%	75%	76%	77%	77%	–
ALS	**94%**	**84%**	**85%**	**85%**	**86%**	**86%**	–

4.5. Robustness checks

To further robustify our results, we carry out the following two exercises. First, we use a different data frequency, namely monthly data. Second, we stick to the daily frequency but use alternative (past-window) sample sizes n, corresponding to one year, two years, five years, and ten years.

To save space, we simply report the main findings here; but the detailed results are also available upon request.

For the monthly data frequency we consider only the most common sample sizes, being $n = 60$ and $n = 120$, that is, five and ten years of past data. Compared to daily data the sample size is very small, however, if we consider more than ten years of past data the investment universe shrinks significantly. In sum, we find similar, but less impressive results for monthly stock returns and factors. Also for monthly data we find evidence that WLS and especially ALS reduce HC3 standard errors, however the reduction is often

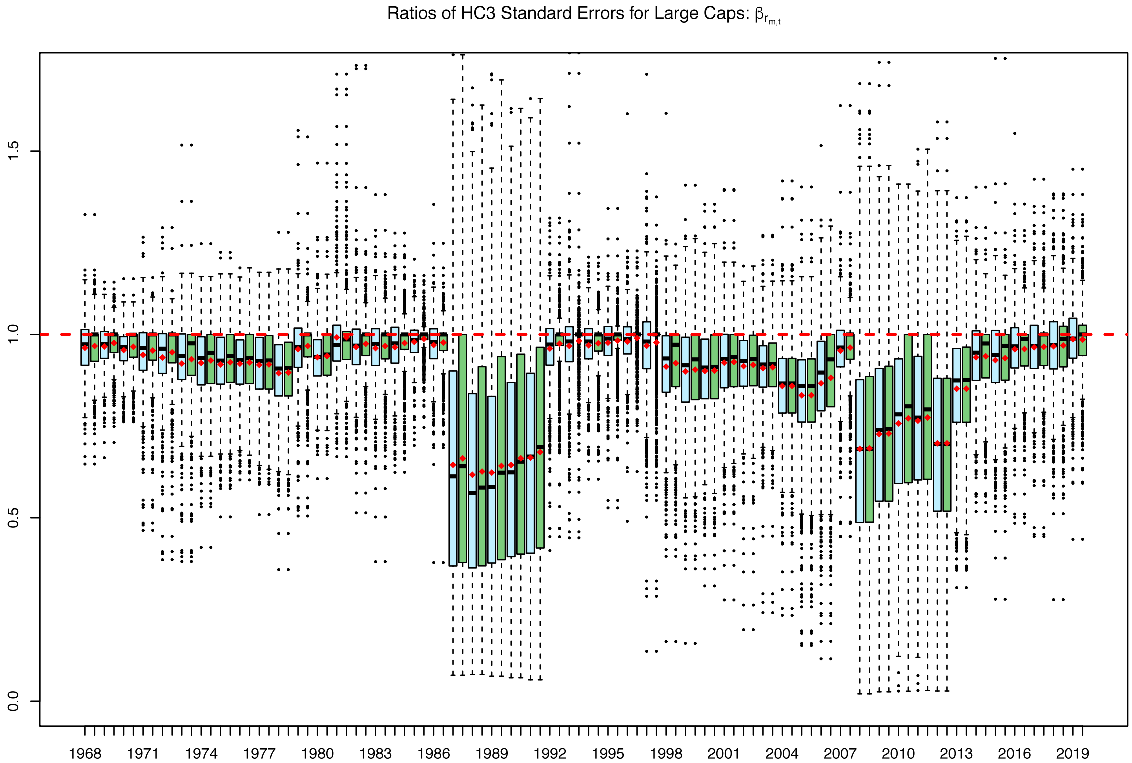

Fig. 9. Five-factor-model boxplots for the market factor β ratios defined in (4.1) for large-cap stocks. For any year, we plot first the WLS/OLS boxplots in **blue**, followed by the ALS/OLS boxplots in **green**. In each box, the bar indicates the median ratio whereas the diamond-shaped symbol indicates the average ratio. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

marginal. Nevertheless, ALS still reduces the HC3 standard errors compared to OLS in 95% of the cases, whereas WLS achieves the same in only 55% of the cases. Consequently, using ALS (or WLS) results in shorter confidence intervals for all investigated coefficients compared to using OLS most of the time, too. However, with monthly data the confidence intervals cannot be reduced as much as for daily data, since conditional heteroskedasticity is less pronounced: On average only in about 15% of the cases the null of homoskedasticity gets rejected.

In terms of different (past-window) sample sizes n, while sticking to the daily frequency, the results are also robust and lead to similar overall conclusions compared to Sections 4.3 and 4.4. Nevertheless, some deviations do exist owing to the fact that the smaller is n, the larger is the eligible investment universe. As the universe increases (by lowering n), most of the additional stocks are micro-caps, so that the percentage of micro-caps increases as well. As a consequence, for smaller sample sizes n, the results become more dispersed and there are more outliers in the boxplots, which leads to somewhat less favorable findings on balance, that is, combined over all caps. On the other hand, within each category — (i) large-cap stocks, (ii) small-cap stocks, and (iii) micro-cap stocks — the benefits of upgrading from OLS to ALS are roughly 'constant' across the different sample sizes n considered; in particular, the largest benefits are obtained throughout for large-cap stocks.

5. Conclusion

In this paper, we show that conditional heteroskedasticity is a common occurrence in CAPM and multi-factor-model regressions and how to carry out improved inference for corresponding regression coefficients. The use of WLS, and especially ALS, has been promoted before by Romano and Wolf (2017). However, we need to extend their theory, since they assume i.i.d. data, which is unrealistic for financial returns. We now demonstrate that the validity of their proposed inference methods based on WLS and ALS continues to hold under a more general set of assumptions that is reasonable for financial factor models; in particular, volatility clustering and (G)ARCH effects can be accommodated.

We run an extensive empirical analysis and find that weighted least squares (WLS) and adaptive least squares (ALS) generally lead to smaller HC standard errors compared to ordinary least squares (OLS). This finding directly translates into shorter confidence intervals and more powerful hypothesis tests. Additionally, we find that ALS consistently outperforms WLS in terms of the percentage of standard errors that are reduced compared to OLS. Note that especially for monthly data, where conditional heteroskedasticity is less pronounced, the flexibility of ALS to choose between WLS and OLS is advantageous. We also find differences with respect to the market capitalization: Generally, the larger the stock, the larger the benefit of using WLS and ALS instead of OLS.

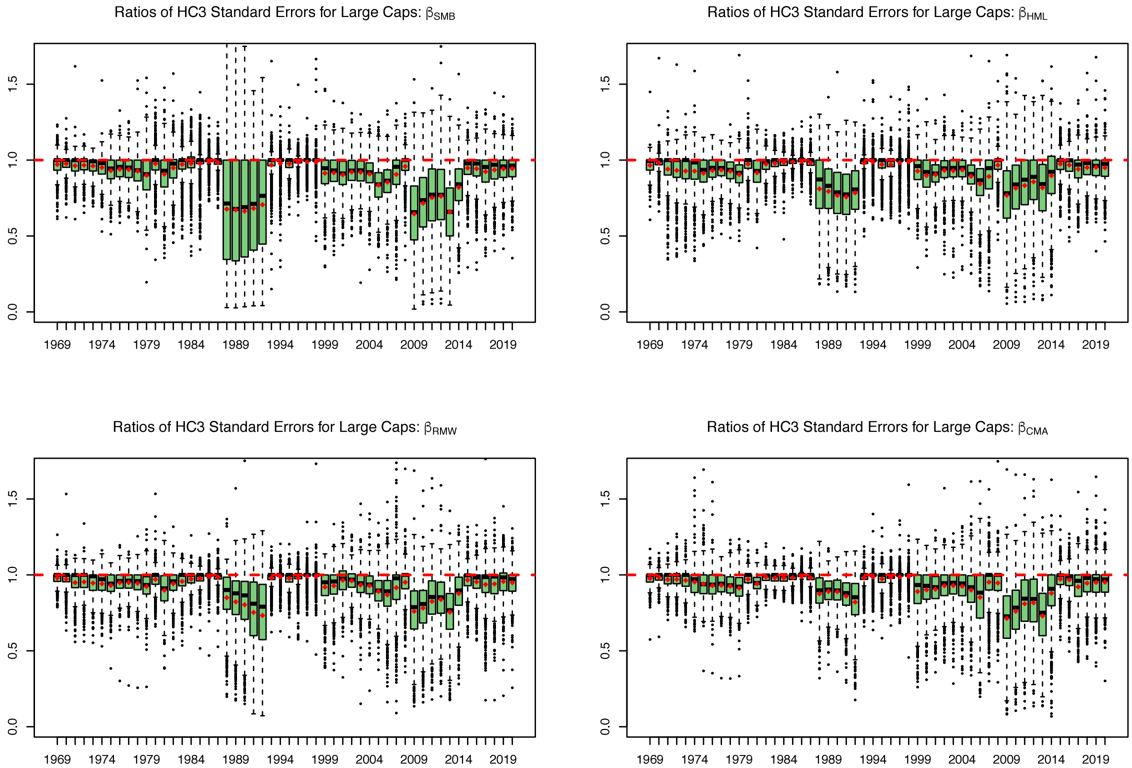

Fig. 10. Five-factor-model boxplots for the SMB, HML, RMW and CMA β's ratios defined in (4.1) for large-cap stocks. Due to readability we plot only the ALS/OLS boxplots in **green**. However, in unreported results we find similar patterns for WLS/OLS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

It can be seen that especially during times of financial turmoil the confidence intervals based on WLS and ALS are much shorter compared to OLS for most of the stocks. This is due to the fact that after a stock market crash or during a financial crisis, volatility clustering is much more pronounced, especially for daily data. For some years, and larger stocks, WLS and ALS can cut standard errors almost in half compared to OLS. Further research could investigate if our findings hold for stocks outside of the US and, more generally, for other asset classes.

To sum up, the still-quite-common practice of using OLS in conjunction with the 'usual' standard errors based on an assumption of conditional homoskedasticity should be abandoned because it generally leads to invalid inference. Using OLS in conjunction with HC standard errors fixes this problem but an even better practice is to use WLS or ALS with HC standard errors. In the end, our specific proposal is to use ALS with HC3 standard errors for the overall best performance.

Last but not least, our methodology should not be used when asset or factor returns display noticeable serial correlation as is, for example, the case in hedge-fund performance evaluation; for example, see Fung et al. (2008) and the references therein.

CRediT authorship contribution statement

Elliot Beck: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. **Gianluca De Nard:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Michael Wolf:** Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Appendix. Proof of Theorem 3.1

To show that the stated results of Romano and Wolf (2017) continue to hold under our more general set of assumption always uses the same method of verification. For this reason, it is enough consider Lemma 3.1 as a typical example.

Fig. 11. Percentages of standard-error ratios (4.1) that lie below one for the five-factor model (1.5) based on large-cap stocks.

Consider the Proof of Lemma 3.1 in Appendix B.1 of Romano and Wolf (2017). Equality (B.1) of course continues to hold and so we are still left to show (B.2) and (B.3).

(B.2) follows by ergodic stationarity, that is, by our more general Assumption (A.2) from the Ergodic Theorem; for example, see Hayashi (2000, p. 101).

To show (B.3) note that

$$X'W^{-1}X = \sum_{t=1}^{n} u_t \quad \text{with} \quad u_t := \frac{x_t \varepsilon_t}{w(x_t)} .$$

It follows from Assumption (A.2) that the sequence $\{u_i\}$ is strictly stationary and ergodic. It follows from Assumption (A.3) that $\{u_t\}$ is a martingale difference sequence; to see this let $z_t := (x_t, \varepsilon_t)'$ and use the "tower property" of conditional expectation, for example, see Williams (1991, Section 9.7):

$$\mathbb{E}(u_t|u_{t-1},\ldots,u_t) = \mathbb{E}(\mathbb{E}(u_t|x_t, z_{t-1},\ldots,z_1)|u_{t-1},\ldots,u_1) .$$

Since

$$\mathbb{E}(u_t|x_t, z_{t-1},\ldots,z_1) = \mathbb{E}\left(\frac{x_t \varepsilon_t}{w(x_t)}|x_t, z_{t-1},\ldots,z_1\right) = \frac{x_t}{w(x_t)}\mathbb{E}(\varepsilon_t|x_t, z_{t-1},\ldots,z_1) = 0 ,$$

where the final equality follows from our more general Assumption (A.3), we have established that $\mathbb{E}(u_t|u_{t-1},\ldots,u_t) = 0$, that is, the MDS property.

The MDS property implies that $\mathbb{E}(u_t) = 0$. The fact that $\mathbb{E}(u_t u_t') = \Omega_{1/w}$ is derived in the identical fashion as in the proof of Romano and Wolf (2017). The result (B.3) then follows by applying the CLT for ergodic stationary MDS; for example, see Hayashi (2000, p. 106). This ends the proof.

The proof of Lemma 3.1 illustrates the same two 'tricks' that are used over and over again also in the proofs of the remaining results. On the one hand, the fact that sample averages converge in probability to population expectations follows from the Ergodic Theorem, as opposed to the strong law of large numbers for i.i.d. sequences; this is where our more general Assumption (A.2) comes into play. On the other hand, the fact that certain (standardized) quantities have a limiting normal distribution follows from the CLT for ergodic stationary MDS, as opposed to the Lindeberg–Lévy CLT for i.i.d. sequences; this is where our more general Assumption (A.3) comes into play.

References

Campbell, J., Lo, A., & MacKinlay, C. (2012). *The Econometrics of Financial Markets* (2nd ed.). Princeton, New Jersey: Princeton University Press.

Carhart, M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, *66*(4), 1047–1108.

Damodaran, A. (2012). *Invetment Valuation*. Hoboken: Wiley.

De Nard, G., Ledoit, O., & Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, *19*(2), 236–257.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, *18*(3), 25–46.

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance*, *63*(4), 1653–1678.

Fama, E. F., & French, K. R. (2015). A five factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*, 1–25.

Fung, W., Hsieh, D. A., Naik, N. Y., & Ramadorai, T. (2008). Hedge funds: Performance, risk, and capital formation. *The Journal of Finance*, *63*(4), 1777–1803.

Green, J., Hand, J. R. M., & Zhang, X. F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, *18*, 692–730.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Hayashi, F. (2000). *Econometrics*. Princeton, New Jersey: Princeton University Press.

Hou, K., Xue, C., & Zhang, L. (2017). *Replicating anomalies*: *Working Paper 23394*, NBER Working Paper Series.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, *47*(1), 13–37.

Long, J. S., & Ervin, L. H. (2000). Using heteroskedasticity consistent standard errors in the linear regression model. *American Statistician*, *54*, 217–224.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, *71*(1), 5–32.

Romano, J. P., & Wolf, M. (2017). Resurrecting weighted least squares. *Journal of Econometrics*, *197*, 1–19.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, *13*(3), 341–360.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442.

Sterchi, M., & Wolf, M. (2017). Weighted least squares and adaptive least squares: Further empirical evidence. In V. Kreinovich, S. Sriboonchitta, & V.-N. Huynh (Eds.), *Robustness in Econometrics* (pp. 135–167). Cham: Springer International Publishing.

Stock, J. H., & Watson, M. M. (2019). *Introduction to Econometrics* (4th ed.). Boston: Pearson.

White, H. L. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica*, *48*, 817–838.

Williams, D. (1991). *Probability with Martingales*. Cambridge: Cambridge University Press.

Wooldridge, J. M. (2012). *Introductory Econometrics* (5th ed.). Mason, Ohio: South-Western.