Capital Share Risk in U.S. Asset Pricing: A Reappraisal

NIKOLAY GOSPODINOV and CESARE ROBOTTI*

May 2020

ABSTRACT

Using long-horizon beta estimates, Lettau, Ludvigson, and Ma (2019) document striking pricing ability and cross-sectional explanatory power for a capital share growth factor across major asset classes. We revisit their findings and show that the spectacular performance of their factor is likely due to the interaction between the lack of identification of the proposed single-factor model and the persistence induced by the overlapping of the data for the purpose of obtaining long-horizon beta estimates. This casts doubts on whether capital share betas are truly priced in the cross-section of asset expected returns.

Keywords: Capital share risk; Long-horizon betas; Model identification; Overlapping data; Small-sample inference.

JEL classification: G12; G20; C12; C14; C15.

*Nikolay Gospodinov is with the Federal Reserve Bank of Atlanta. Cesare Robotti is with the University of Warwick. We thank Martin Lettau, Sydney Ludvigson, and Sai Ma for sharing their data and codes and for providing some clarifications. The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Lettau, Ludvigson, and Ma (LLM, 2019) provide evidence that the capital share growth of aggregate income carries a positive and statistically significant risk premium for a wide range of assets. The main results are based on long-horizon betas that are obtained by regressing H-period compounded test asset returns on the H-period growth of the capital share factor. The authors argue that this H-period aggregation can mitigate the effect of measurement error in the data as well as the impact of model misspecification arising from omission of additional risk factors. The estimated H-horizon betas are then shown to be useful for explaining one-period expected return premia.

We revisit some of the findings in the paper, and in particular the empirical support for some implicit assumptions that underpin the validity of the statistical inference. More specifically, the validity of the statistical inference depends crucially on the existence of nonzero covariances between the asset returns and the factor. As stated in Pukthuanthong, Roll, and Subrahmanyam (2019), one of the fundamental attributes of a true risk factor is that "its variations induce changes in asset prices." The earlier literature refers to a factor that lacks this attribute as a "useless" or "spurious" factor. (See Kan and Zhan 1999; Gospodinov, Kan, and Robotti 2017.) Statistically, this factor attribute amounts to an identification condition for the risk premia parameters. Subsequently, if the factor does not satisfy this decision and the inference procedure is not robust to this feature, the results for the risk premia could be highly misleading. Below, we subject this maintained assumption to more scrutiny given its central role in ensuring reliability of the empirical results. The persistence induced by constructing H-period returns and factors as well as the large number of test assets relative to the number of time series observations can further distort inference in finite samples and properties of standard diagnostic tools.

The main results in LLM can be summarized as follows. First, LLM report strong evidence of pricing for their capital share factor. The pricing ability of their single-factor model is not limited to equity portfolios, but it also extends to other asset classes such as corporate bonds, sovereign bonds, options, and credit default swaps (CDS). Moreover, the ordinary least squares (OLS) cross-sectional R^2s are generally very high, thus indicating that the betas line up well with the one-period expected returns. Finally, LLM argue that their proposed factor is not spurious.

The central findings in LLM for different equity portfolios are presented in a graphical form in Figure 1.[1]

[1] We focus on LLM's equity portfolios in the paper, whereas the analysis for nonequity asset classes is in the

Figure 1 about here

While LLM report results only for horizons $H = 4$ and 8, Figure 1 presents the capital share risk premium estimate and its 95% bootstrap confidence interval for $H = 1, 2, \ldots, 16$.[2] We expect this more comprehensive range to facilitate our understanding of the sampling behavior of the risk premium estimator as H increases. Figure 1 illustrates the strong significance results and tight confidence intervals for $H = 4$ and 8 reported in LLM. The tight confidence intervals are particularly striking. As a general pattern, the length of the 95% confidence intervals decreases steadily as H increases. For example, for all equities (85 equity portfolio returns), the 95% bootstrap confidence intervals for $H = 1$, 8, and 16 are [0.08, 1.08], [0.49, 0.65], and [0.28, 0.38], respectively. Imposing, in the bootstrap samples, the null that the parameter is zero results in a 95% confidence interval of [−0.00, 0.00] for $H = 8$ (as reported in Table IA.VII of LLM's Internet Appendix, which is also available on the *Journal of Finance* website).[3] This is accompanied by remarkably high cross-sectional regression R^2s with values of 0.95 for options ($H = 4$), 0.94 for CDS ($H = 8$), and 0.89 for bonds ($H = 8$). For context, the number of time series observations used for computing these R^2s is 98 (18 test assets), 38 (20 test assets), and 139 (20 test assets), respectively. Given the relatively small number of time series observations – which is further exacerbated by the overlapping of the data – and the large number of test portfolios, the purported accuracy of the statistical inference raises some red flags about the validity of these confidence intervals. In what follows, we attempt to understand what causes the tightness of the bootstrap confidence intervals and whether LLM's main empirical results continue to hold under additional scrutiny.

The rest of the paper is organized as follows. Section I introduces the notation and presents the main results of our analysis. In this section, we analyze the relationship between model identification and the candidate factor's pricing ability. In addition, we explain why the OLS cross-sectional R^2 is very high and explore the statistical properties of the various tests. Section II summarizes our main findings. Additional material is provided in the Internet Appendix.

I. Main Results

First, we introduce the notation and describe the data used in the empirical analysis. Next, we

Internet Appendix.

[2]Figure 1 is based on LLM's code that is available on the *Journal of Finance* website.

[3]The 95% confidence intervals for several other test assets in Table IA.VII are also [−0.00, 0.00].

present our main results. Simulation results supporting the use of our tests are deferred to the end of the section and to the Internet Appendix so that we can get right to the empirical analysis.

A. Notation

The notation that we adopt is the same as in LLM but we introduce it again for completeness and ease of exposition. Let T denote the time series sample size. The vector $\mathbf{R}_{t+H,t} = [R_{1,t+H,t}, \ldots, R_{N,t+H,t}]'$ denotes an $N \times 1$ vector of H-period gross returns on the test assets between time t and $t+H$ for $t = 1, \ldots, T-H$. Its j-th element ($j = 1, \ldots, N$) is defined as $R_{j,t+H,t} = \prod_{h=1}^{H} R_{j,t+h}$, where $R_{j,t+h}$ is the j-th gross one-period (nonoverlapping) return between time $t+h-1$ and time $t+h$. Similarly, the H-period gross risk-free rate, $R_{f,t+H,t}$, is defined as $R_{f,t+H,t} = \prod_{h=1}^{H} R_{f,t+h}$, where $R_{f,t+h}$ is the gross one-period (nonoverlapping) risk-free rate between time $t+h-1$ and time $t+h$. Let $\mathbf{R}^e_{t+H,t} = \mathbf{R}_{t+H,t} - \mathbf{1}_N R_{f,t+H,t}$ be the corresponding vector of excess returns with a typical element $R^e_{j,t+H,t} = \prod_{h=1}^{H} R_{j,t+h} - \prod_{h=1}^{H} R_{f,t+h}$, where $\mathbf{1}_N$ denotes an N-vector of ones. The H-period risk factor is defined as the capital share at time $t+H$ over the capital share at time t, $f_{t+H,t} = KS_{t+H}/KS_t$, where $KS_t = 1 - LS_t$ and LS_t is the labor share of the nonfarm business sector. When $H = 1$, the excess returns and the factor are one-period (nonoverlapping) returns and for notational convenience we denote them by $\mathbf{R}^e_{t+1} = [R_{1,t+1} - R_{f,t+1}, \ldots, R_{N,t+1} - R_{f,t+1}]'$, with $\boldsymbol{\mu}_R \equiv \mathbb{E}[\mathbf{R}^e_{t+1}]$, and $f_{t+1} = KS_{t+1}/KS_t$.

The two-pass procedure adopted in LLM can be described as follows. The H-period vector of risk exposures, denoted by $\boldsymbol{\beta}_H = [\beta_{1,H}, \ldots, \beta_{N,H}]'$, is estimated from a time series regression of H-period excess returns on the H-period factor:

$$\mathbf{R}^e_{t+H,t} = \boldsymbol{\alpha} + \boldsymbol{\beta}_H f_{t+H,t} + \boldsymbol{\epsilon}_{t+H,t}, \tag{1}$$

where $\boldsymbol{\epsilon}_{t+H,t}$ is an N-dimensional vector of error terms. Given the possible presence of measurement error in the data and the potential misspecification arising from omitting relevant but latent factors, LLM conjecture that the estimated (by OLS) longer-horizon risk exposures, $\hat{\boldsymbol{\beta}}_H$, are likely to be less biased than the one-period beta estimates for the true one-period betas.[4] Let $\hat{\boldsymbol{\mu}}_R$ be the sample equivalent of $\boldsymbol{\mu}_R$. Then, LLM run a cross-sectional regression of $\hat{\boldsymbol{\mu}}_R$ on the H-period risk exposure

[4] In Section III of their Internet Appendix, LLM build a numerical example, which they view as a possibility result, in order to show that it is plausible that the proposed long-horizon beta estimates are more unbiased than the one-period beta estimates for the true one-period betas, at least at some particular horizon H. The main problem with their numerical example is that it leads to the counterfactual result that the return volatility is smaller than the KS factor volatility, whereas, in the data, the returns are one order of magnitude more volatile than the KS factor.

estimates as follows:

$$\hat{\boldsymbol{\mu}}_R = \mathbf{1}_N\lambda_0 + \hat{\boldsymbol{\beta}}_H\lambda_H + \mathbf{e}, \tag{2}$$

where λ_0 is the zero-beta rate, λ_H is the risk premium parameter for the capital share factor, and $\mathbf{e}$ is an N-vector of model pricing errors. Let $\mathbf{X} = [\mathbf{1}_N,\ \boldsymbol{\beta}_H]$ and $\hat{\mathbf{X}} = [\mathbf{1}_N,\ \hat{\boldsymbol{\beta}}_H]$. The OLS estimate of $\boldsymbol{\lambda}_H \equiv [\lambda_0, \lambda_H]' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mu}_R$ is given by

$$\hat{\boldsymbol{\lambda}}_H \equiv \begin{bmatrix} \hat{\lambda}_0 \\ \hat{\lambda}_H \end{bmatrix} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\hat{\boldsymbol{\mu}}_R, \tag{3}$$

and the (adjusted) OLS cross-sectional R^2 metric is

$$R^2_{adj} = 1 - \left(\frac{N-1}{N-2}\right)\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\hat{\mathbf{e}}_0{}'\hat{\mathbf{e}}_0}, \tag{4}$$

where $\hat{\mathbf{e}} = \hat{\boldsymbol{\mu}}_R - \hat{\mathbf{X}}\hat{\boldsymbol{\lambda}}_H$, $\hat{\mathbf{e}}_0 = [\mathbf{I}_N - \mathbf{1}_N(\mathbf{1}_N'\mathbf{1}_N)^{-1}\mathbf{1}_N']\hat{\boldsymbol{\mu}}_R$, and $\mathbf{I}_N$ is an $N \times N$ identity matrix.

Before we proceed with the empirical arguments, we would like to stress two key characteristics of this theoretical setup. First, $\boldsymbol{\lambda}_H$ is defined if $\mathbf{X} = [\mathbf{1}_N,\ \boldsymbol{\beta}_H]$ is of full column rank. Since this is a single factor model ($\boldsymbol{\beta}_H$ is an $N \times 1$ vector), this identification condition can be violated if $\boldsymbol{\beta}_H = \mathbf{0}_N$ or $\boldsymbol{\beta}_H$ is proportional to $\mathbf{1}_N$, where $\mathbf{0}_N$ is an N-vector of zeros. We will discuss these two possibilities later. Second, note that the estimate of interest, $\hat{\lambda}_H$, is unchanged if we work with the gross returns $\mathbf{R}_{t+H,t}$ instead of the excess returns $\mathbf{R}^e_{t+H,t}$. The log transformations of test asset returns and factor can then be expressed as H-period overlapping (telescoping) sums of one-period returns or growth rates: $\ln(R_{j,t+H,t}) = \sum_{h=1}^{H}\ln(R_{j,t+h})$ and $\ln(f_{t+H,t}) = \sum_{h=1}^{H}\triangle\ln(f_{t+h})$, where $\triangle\ln(f_{t+h}) = \ln(KS_{t+h}) - \ln(KS_{t+h-1})$. It is well known that this overlapping structure induces persistence in the H-period series and this persistence becomes stronger (even bordering nonstationarity) as H increases. The interaction of these two features (rank failure and persistence) will play a crucial role in the following analysis.

B. Data

We use LLM's dataset that is available on the *Journal of Finance* website. The return data are from Kenneth French's data library and consist of the monthly value-weighted returns on the (i) 25 size and book-to-market sorted portfolios (Size/BM, $N = 25$); (ii) 10 long-run reversal portfolios (REV, $N = 10$); (iii) 25 size and operating profitability sorted portfolios (Size/OP, $N = 25$); and (iv) 25 size and investment sorted portfolios (Size/INV, $N = 25$). The data are compounded to a quarterly frequency and the one-month T-Bill rate is used as the risk-free rate in forming the excess returns on the test assets. The empirical analysis is carried out by using the four sets of equity portfolios from above individually as well as jointly (All Equities, $N = 85$). The capital share factor of LLM is based on the data compiled by the Bureau of Labor Statistics, and it is measured on a quarterly basis.

LLM's sample period runs from 1963:Q3 to 2013:Q4. However, the beginning date used by LLM in their empirical analysis (before overlapping) is not 1963:Q3. For some reasons, LLM estimate the first-stage betas based on data from 1964:Q1 to 2013:Q4, while the second-stage lambdas are estimated by using one-period returns over the 1963:Q4 – 2013:Q4 sample period. To be consistent with LLM, we report the results for their sample choice. The results for the full sample starting in 1963:Q3 are overall similar to the ones reported in the paper and are available from the authors upon request.

To keep the subsequent analysis in the relevant context, Figure 2 plots the H-period Size/BM excess returns and H-period capital share factor for $H = 1$, 4, and 8.

Figure 2 about here

Despite the difficulties in drawing any strong conclusions from just plotting the data, Figure 2 clearly illustrates the higher persistence of the variables as the amount of compounding increases. While it is obvious that this compounding induces some "structure" in the variables, the statistical analysis should reveal if any potential co-movement is genuine or arises purely from commonality in the H-period overlapping. Finally, it is worth pointing out that the variability of returns greatly exceeds the volatility of the factor as the "signal-to-noise" ratio decreases further with the horizon H.

C. Factor Loadings and Tests for Rank Failure

The crucial identification condition is that the matrix $\mathbf{X}$ is of full column rank. Before we subject this identification condition to formal statistical testing, we glean some information from the sampling behavior of $\hat{\boldsymbol{\beta}}_H$ as H increases.

Figure 3 about here

Figure 3 plots the minimum, median, and maximum values of the N-vector $\hat{\boldsymbol{\beta}}_H$ for $H = 1, \ldots, 16$ for the different equity portfolios. The figure clearly illustrates that the volatility and spread across assets increase with H. Provided that the identification condition is satisfied, the larger variability of $\hat{\boldsymbol{\beta}}_H$, which are the regressors in the second-pass regression, would result in higher precision for $\hat{\lambda}_H$.

It should be noted that characterizing the sampling behavior of $\hat{\boldsymbol{\beta}}_H$ for $H > 1$ can be quite challenging. In the general case, the sampling distribution of $\hat{\boldsymbol{\beta}}_H$ depends on three quantities: T, N, and H. When $N = 1$, Valkanov (2003) derives the asymptotic distribution of the scalar $\hat{\beta}_H$ under the assumption that the horizon H increases linearly with the sample size T. This is a convenient statistical device that provides a more accurate approximation of the finite-sample distribution of $\hat{\beta}_H$ when H is a non-trivial fraction of T. Importantly, Valkanov (2003) demonstrates that the H-period overlapping helps in reducing the noise component and improves the estimation accuracy (as T gets large).[5] At the same time, H-period overlapping reduces the effective number of time series observations. In other words, overlapping data trims the independent information in the sample, even when T is large. (See Richardson and Stock 1989.) The independent information in the sample is further reduced when $N > 1$, which is the case considered here. Since in LLM's setting the two-pass procedure can be rewritten as a moment condition problem with $3N$ moment conditions, the effective number of time series observations per moment condition (test asset) could be quite limited for large N (and H). This makes the bootstrap a preferable method for conducting inference for this complex, small-sample problem.

As stated above, the validity of the standard methods of inference, including the parametric bootstrap adopted by LLM, depends crucially on the condition that $\mathbf{X}$ is of full column rank. Given the small T, large N, and possibly large H, standard asymptotic joint beta and rank tests would provide an extremely inaccurate approximation of the sampling distributions of these statistics. (Simulation experiments that illustrate this point – with empirical rejection rates of these tests as large as 100% at the 5% significance level – can be found in the Internet Appendix.) For this reason, we employ bootstrap versions of these tests that provide a substantial size improvement over their asymptotic counterparts. (The tests, bootstrap procedure, and empirical rejection rates

[5] Valkanov (2003) considers a one-period factor with a near-unit root which drives the super-consistency of the factor estimates in his analysis. In our setup, the factor (KS growth, consumption growth, market factor) is stationary. In this case, the beta estimates are inconsistent as H increases linearly with the sample size (the results are available from the authors upon request).

are reported in the Internet Appendix.) More specifically we construct bootstrap tests of the hypotheses $H_0 : \boldsymbol{\beta}_H = \mathbf{0}_N$ and $H_0 : \text{rank}(\mathbf{X}) = 1$.[6] Figure 4 presents the bootstrap p-values of these tests.

Figure 4 about here

Based on a 5% significance level, we can never reject the two null hypotheses stated above for all H. The two tests typically deliver p-values that are very close to each other and, in the "All Equities" case, these p-values are extremely large (0.7 and above) for all H. It should also be noted that for $H = 1$ (the non-overlapping, one-period beta scenario), we cannot reject the null of lack of identification. Our results suggest that if a factor is spurious in one-period regressions, it is quite unlikely that it will become useful by overlapping the data and running long-horizon regressions. Put differently, in a spurious regressor setting, the spread and variability of the beta estimates will typically increase with the horizon, but this increased variability will also negatively affect the power of the joint beta and rank tests.

It is important to stress that significance tests on individual betas or beta spreads (as in Table VII in LLM), while informative for some purposes, would not address the issue of identification failure. Statistical significance of individual betas or spreads – as it is indeed the case in this application – is consistent with the possibility of identification failure, i.e., that all betas are jointly statistically indistinguishable from a vector of zeros. This occurs when the betas across the different test assets are highly correlated, as reflected in the large off-diagonal elements of the covariance matrix of $\hat{\boldsymbol{\beta}}_H$. An informal diagnostic to gauge how far is the covariance matrix of $\hat{\boldsymbol{\beta}}_H$, $\mathbf{V}_{\hat{\boldsymbol{\beta}}_H}$, from a diagonal matrix with $\text{diag}(\mathbf{V}_{\hat{\boldsymbol{\beta}}_H})$ on the main diagonal, which is used for the individual tests, is the ratio between the largest eigenvalue of these two respective matrices. With 1 being the value of this ratio when the two matrices have identical largest eigenvalues, the ratio for the 25 size and book-to-market equity portfolios ranges between 7.0 and 9.4 for $H = 1, \ldots, 8$. This suggests that significance tests on the individual betas or beta spreads would not be appropriate for evaluating the validity of the identification condition.

[6]The joint beta and rank tests could, in principle, deliver different answers. With a constant in the second-pass regression, rank deficiency can occur when $\boldsymbol{\beta}_H$ is a zero vector as well as when $\boldsymbol{\beta}_H$ is a constant vector. For this reason, we report both tests in Figure 4. The Internet Appendix reports a numerical illustration of this observation which is also discussed later in this section.

In Section III.D and Table VII, LLM recognize that the validity of their empirical results crucially hinges upon their single-factor model being well-identified. Overall, they argue that the considered beta spreads are large and significantly different from zero, which implies that their factor is not spurious. However, it should be noted that LLM reach this conclusion by only focusing on the spread in betas between the highest and lowest average (one-period) return portfolio for each portfolio group. By considering only one particular beta spread, LLM completely disregard the information content in all the other assets in a given portfolio group. This begs the question of how, if we were to adopt LLM's strategy in a setting where asset returns are highly correlated with each other (as it is the case here), one would then combine the inference results for the individual spreads within a given portfolio group in order to determine whether the betas are jointly zero (constant) or not. This is precisely the type of questions that joint beta and rank tests try to address. Clearly, individual tests, such as the ones in LLM, would deliver qualitatively similar conclusions to the ones of our joint tests if the off-diagonal elements of the covariance matrix of $\hat{\boldsymbol{\beta}}_H$ were to be small. But for all asset classes considered by LLM, these off-diagonal elements are large, as the eigenvalue-based statistics mentioned in the previous paragraph indicate. This is the driver of the substantial differences between our and their results.

Is this type of behavior specific to the KS factor? How do the rank failure and H-period overlapping affect other factors? Two factors that could help us to gain further intuition are the consumption factor and the market factor supplied in LLM's dataset.[7] The full set of results is in our Internet Appendix and can be summarized as follows. Despite that the consumption factor is characterized by a very different dynamics compared to the KS factor, the confidence intervals for λ_H are also extremely tight, although they signal less significance relative to the KS factor case. The joint beta and rank tests strongly point to rank failure also for consumption. On the other hand, the market factor is not expected to be a spurious factor. Interestingly, despite its high correlation with the equity returns, LLM's bootstrap procedure finds little pricing evidence for the market factor with much wider confidence intervals. Furthermore, the rank condition for the H-period CAPM also appears to be compromised but now not because $\boldsymbol{\beta}_H = \mathbf{0}_N$ but because $\hat{\boldsymbol{\beta}}_H$ cannot be statistically distinguished from a column of ones. (The interested readers are invited

[7]We follow LLM and define the H-period consumption factor as C_{t+H}/C_t, where C is the expenditures on non-durables and services (excluding shoes and clothing). Because this is not a log growth rate, the factor exhibits a mild trending behavior for large H. The market factor is the excess return on the value-weighted NYSE-AMEX-NASDAQ stock market index.

to consult the Internet Appendix for more details.)

In summary, the results in this section suggest that LLM's factor is spurious regardless of the chosen horizon. This observation is important because, in the next subsection, it will help us rationalize the tight confidence intervals for the risk premia and the large cross-sectional R^2 values documented by LLM.

D. Risk Premia and Cross-Sectional R^2

In what follows, we undertake some numerical experiments aimed at showing that significant risk premia and high cross-sectional R^2 statistics can arise even when the factor is spurious, i.e., it exhibits no pricing ability for a given set of test asset returns. Detecting the underlying lack of pricing ability, however, may not be straightforward as it can be obscured by high persistence, induced by H-period compounding, and the small number of effective time series observations. Below we illustrate how the standard analysis can be distorted by these characteristics of the data.

In our first experiment, we generate a spurious factor – a factor that is independent of the test asset returns – which, by construction, should carry a zero risk premium and no cross-sectional explanatory power. In order to accomplish this and preserve the salient features of the data, we start by approximating the capital share as an autoregressive (AR) of order one, AR(1), process:

$$KS_{t+1} = \delta + \rho KS_t + \varepsilon_{t+1}, \tag{5}$$

where ε_{t+1} is a mean-zero error term with variance σ^2 and $\rho < 1$.[8] Let $\hat{\sigma}^2$ denote the sample estimate of σ^2. We consider two designs: (i) ε_{t+1} is generated as $N(0, \hat{\sigma}^2)$ and is independent of $[R_{1,t+1} - R_{f,t+1}, \ldots, R_{N,t+1} - R_{f,t+1}]'$; and (ii) ε_{t+1} is generated as $N(0, \hat{\sigma}^2)$ but with covariances with $[R_{1,t+1} - R_{f,t+1}, \ldots, R_{N,t+1} - R_{f,t+1}]'$ that are calibrated to the sample covariances from the data.[9] Thus, case (i) corresponds to the case of a spurious factor. After drawing a sample path $\{\varepsilon^{\circ}_{t+1}\}$ either under (i) or (ii), we construct a simulated capital share series $KS^{\circ}_{t+1} = \hat{\delta} + \hat{\rho} KS^{\circ}_t + \varepsilon^{\circ}_{t+1}$ for

[8] Higher-order AR processes do not appear to provide better approximations. An alternative model specification involves modeling the deterministic component with a break in trend in 2001:Q2 (see, for example, Elsby, Hobijn, and Sahin 2013), while the stochastic component is still an AR(1) process. Finally, we also accounted for the time-varying volatility in the errors by means of an AR-generalized autoregressive conditional heteroskedasticity model. We experimented with all these alternative specifications but only report our findings for the pure AR(1) model since the results are very similar.

[9] As an alternative to case (ii), we also generated data using the full parametric setup of the two-pass regression framework with the sample estimates $\hat{\boldsymbol{\lambda}}_H$ treated as true values. The results are essentially the same as those for simulation design (ii) and are not reported.

some initial value KS_1,[10] where $\hat{\delta}$ and $\hat{\rho}$ denote the OLS estimates of δ and ρ, respectively. We can then use the simulated capital share process to construct the H-period factor $f^{\circ}_{t+H,t} = KS^{\circ}_{t+H}/KS^{\circ}_{t}$.

As pointed out in the introduction, the cross-sectional R^2 values reported by LLM are very high, ranging from 0.62 to 0.86 for equities at horizon $H = 8$. Gospodinov, Kan, and Robotti (2019) argue, in the context of invariant estimators, that cross-sectional R^2 statistics close to one could well be a symptom of spurious rather than genuine explanatory power. In order to assess whether it is possible to generate the magnitude and shape of the R^2 profiles observed in data with a spurious factor, we replace the actual factor $f_{t+H,t}$ with the simulated factor $f^{\circ}_{t+H,t}$ for case (i). We then generate 10,000 factor sample paths,[11] re-estimate the parameters of the two-pass regression model and compute the resulting R^2 measures. Finally, we take the $75th$ percentile of the empirical distribution of the simulated R^2s at $H = 8$ and the R^2s that are associated with this value at the remaining H horizons. Figure 5 plots the sample and the simulated R^2 profiles for $H = 1, \ldots, 16$.

Figure 5 about here

The plots show that the sample cross-sectional R^2s for nonoverlapping data ($H = 1$) are around zero. Combined with our previous graphs, if the researcher were to investigate the pricing performance of the capital share factor in the standard setup with nonoverlapping data, no evidence of pricing ability would be found. The R^2s start to increase consistently with the horizon and peak (at high levels) between $H = 6$ and $H = 10$. It is tempting to conclude that this lends support to the argument of strengthened signal and reduced misspecification at longer horizons. But Figure 5 reveals that this sample pattern is also consistent with a spurious factor that is independent of the test asset returns. The "commonality" between the factor and the returns can be traced back to the common persistent pattern from H-period overlapping, amplified by the small effective time series sample size. The results are even more pronounced for the nonequity asset classes where the time series span is substantially shorter. (See the Internet Appendix.) Thus, relying on standard measures of significance and goodness-of-fit in this setup could be highly misleading.

[10] As for the initial value, we draw randomly an observation from the sample and employ it to start the recursion. We also experimented with the first observation in the sample and found the difference in results to be negligible.

[11] Here, we keep test asset returns fixed at their sample values. This way we do not need to assume a parametric data generating process for returns that replicates well all of their sample properties. In unreported results, we also simulate a panel of return series and find that the difference in results is negligible.

Next, we present some simulation evidence on the coverage probabilities of the bootstrap procedure used by LLM for horizons $H = 1$, 4, and 8. Specifically, we consider cases (i) and (ii) above in Panels A and B of Table I, respectively. For the sake of comparison, we also develop a nonparametric block bootstrap method that, unlike LLM's bootstrap, is agnostic to the parametric structure of the model and exhibits robustness to model misspecification and poor identification. The details for this method are described in the Internet Appendix.

Table I about here

In Panel A (spurious factor case), we report the coverage probabilities of the 95% and 90% confidence intervals based on LLM's bootstrap method[12] (denoted LLMP in the table) and the nonparametric bootstrap (NP). Regardless of the chosen horizon H, LLM's bootstrap confidence interval drastically undercovers the pseudo-true value of the risk premium parameter (which is set equal to zero for the spurious factor case). For example, for 95% confidence intervals, the coverage probabilities are 34.6%, 43.0%, and 40.4% for $H = 1$, 4, and 8, respectively. In contrast, the coverage probabilities of the confidence intervals based on the nonparametric bootstrap are much closer to their theoretical counterparts. For 95% confidence intervals, these coverage probabilities are 98.5%, 99.9%, and 100.0% for $H = 1$, 4, and 8, respectively.[13] In the same panel, we also report the median length of the confidence intervals for the two bootstrap methods at various horizons H. The parametric bootstrap method of LLM delivers much tighter confidence intervals relative to the nonparametric bootstrap. For example, for $H = 8$, the median length of LLM's bootstrap confidence interval is 0.927 while the one for the nonparametric bootstrap is 1.995. So, Panel A emphasizes the vastly different results across bootstrap methods when the factor is completely spurious.[14]

Panel B of Table I (capital share factor case) replaces the spurious factor with a factor that mimics the dynamics and return correlation properties of the capital share factor proposed by LLM.

[12]For all of the results in the paper, we use the original code for LLM's bootstrap which is available from the *Journal of Finance* website. In the Internet Appendix, we provide a modified version of LLM's bootstrap that corrects the type of resampling in the first and second pass, while retaining the remaining structure of LLM's bootstrap procedure. Simulation results for this modified method (with improved properties over the original LLM's method) are available from the authors upon request.

[13]In unreported simulations, the coverage probabilities for the nonparametric bootstrap at $H = 1$ approach the nominal level as the sample size increases. For example, for $T = 1000$, the coverage probabilities of the 95% and 90% nonparametric bootstrap confidence intervals are 95.8% and 91.4% for the spurious factor case (i), and 95.2% and 90.5% for the capital share factor case (ii), respectively.

[14]In principle, when the factor is useful, the parametric and nonparametric bootstrap methods should exhibit similar coverage properties.

The use of this factor delivers results that are very similar to the ones for the spurious factor case in Panel A. This should not come as a surprise since based on the tests and descriptive statistics in Figures 4 and 5, we cannot reject the null hypothesis that LLM's factor is spurious. Overall, Table I indicates that the coverage properties of LLM's confidence interval methodology are very poor in a (nearly) spurious factor setting.

Given the good (albeit conservative) coverage properties of the nonparametric bootstrap confidence intervals when the factor is spurious or nearly spurious, Figure 6 plots these 95% confidence intervals for the actual sample.[15]

Figure 6 about here

For all equity groups and compounding horizons H, the 95% confidence intervals based on the nonparametric bootstrap contain the zero value, thus indicating that the capital share factor of LLM is not priced. (The results are even stronger for nonequity asset classes, as emphasized in the Internet Appendix.) This is in stark contrast with Figure 1, the core of LLM's empirical analysis. The differences between Figures 1 and 6 can be traced back to the spuriousness of the capital share factor and the completely different coverage probabilities and confidence interval lengths of the two bootstrap methods in a spurious factor setting. In summary, our analysis lends little or no support to LLM's pricing conclusions. Overlapping the data does not necessarily lead to a higher signal to noise ratio, and if a factor is spurious before compounding it may as well remain so after compounding.

II. Concluding Remarks

The evidence presented in this paper does not lend support to the findings of stark statistical significance and explanatory power reported in LLM. The source of this reappraisal of the original results in LLM stems from lack of identification of the risk premium parameter that is partially obscured by the persistence induced by H-period compounding of the data. Poor or incomplete identification renders the statistical inference highly non-standard and inference procedures, asymptotic or bootstrap, that explicitly or implicitly maintain the identification assumption would likely

[15]Establishing the asymptotic validity of this nonparametric bootstrap for the two-pass (H-period) regression framework is beyond the scope of this paper.

be misleading by favoring near- or even completely uninformative factors. While it is possible that transformations of the original data may strengthen the underlying signal, the statistical method should account for the possibility of spurious commonality arising from overlapping data over long horizons. Finally, we highlight the need for additional caution due to the reduced number of effective time series observations per test asset for longer horizons and the large number of test assets.

We would like to conclude with a few general observations. Over the last decade, there has been a proliferation of proposed factors for pricing asset returns. More recently, several competing one- or two-factor models have been advanced that individually have been shown to offer strong explanatory power across asset classes. Conceptually, this presents some challenges to reconciling these results as the candidate factors carry disparate economic content. Empirically, these studies evaluate the pricing performance of the various factors using different statistical and inference setups and methods that not always account for the uncertain properties of data and models. Often, when subjected to more rigorous and robust testing, the evidence of improved pricing performance of these parsimonious factor models largely disappears.[16]

This calls for a unified framework for factor evaluation that explicitly recognizes model incompleteness, possibility of under-identification, effects of persistence and limited sample information, choice of test assets, etc. on the inference outcomes. The methodological progress in the direction of ensuring robustness of statistical inference in asset pricing models has been incremental but steady. Naturally, this trades off efficiency and (often spurious) statistical significance for increased reliability (albeit subdued sharpness) of the results. Despite some encouraging signs, the empirical literature is yet to fully embrace a more general approach to incorporating recent methodological advances in statistical inference. The recent economic and financial developments are another attestation for the need of model resilience to adverse shocks and heightened uncertainty.

[16]For example, Adrian, Etula, and Muir (2014) and He, Kelly, and Manela (2017) propose one- and two-factor models based on a financial intermediary risk narrative, whereas Lettau, Maggiori, and Weber (2014) advocate using the downside-beta CAPM to price various asset classes. Using a unifying framework, Gospodinov and Robotti (2020) revisit these results and show that robust evidence for factor pricing across asset classes remains elusive. Similarly, Kleibergen and Zhan (2020) raise concerns about the reliability of traditional asset pricing tests with unspanned risk factors.

REFERENCES

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial intermediaries and the cross-section of asset returns, *Journal of Finance* 69, 2557–2596.

Elsby, Michael W. L., Bart Hobijn, and Aysegul Sahin, 2013, The decline of the U.S. labor share, *Brookings Papers on Economic Activity* 2, 1–52.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2017, Spurious inference in reduced-rank asset-pricing models, *Econometrica* 85, 1613–1628.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2019, Too good to be true? Fallacies in evaluating risk factor models, *Journal of Financial Economics* 132, 451–471.

Gospodinov, Nikolay, and Cesare Robotti, 2020, Common pricing across asset classes: Empirical evidence revisited, *Journal of Financial Economics*, forthcoming.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

Kan, Raymond, and Chu Zhang, 1999, Two-pass tests of asset pricing models with useless factors, *Journal of Finance* 54, 203–235.

Kleibergen, Frank, and Zhaoguo Zhan, 2020, Robust inference for consumption-based asset pricing, *Journal of Finance* 75, 507–550.

Lettau, Martin, Matteo Maggiori, and Michael Weber, 2014, Conditional risk premia in currency markets and other asset classes, *Journal of Financial Economics* 114, 197–225.

Lettau, Martin, Sydney C. Ludvigson, and Sai Ma, 2019, Capital share risk in U.S. asset pricing, *Journal of Finance* 74, 1753–1792.

Pukthuanthong, Kuntara, Richard Roll, and Avanidhar Subrahmanyam, 2019, A protocol for factor identification, *Review of Financial Studies* 32, 1573–1607.

Richardson, Matthew, and James H. Stock, 1989, Drawing inferences from statistics based on multiyear asset returns, *Journal of Financial Economics* 25, 323–348.

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

Table I
Coverage Probabilities and Confidence Interval Length

The table presents Monte Carlo simulation results for coverage probabilities and median confidence interval length (in square brackets) for two methods: (i) The parametric bootstrap of LLM (LLMP); and (ii) The nonparametric bootstrap (NP) described in the Internet Appendix. The number of simulated factor paths is 10,000. The test returns are kept fixed at their observed values across the simulations. For each of the 10,000 Monte Carlo iterations, the number of bootstrap replications is set equal to 399. H indicates the horizon in quarters over which returns and capital share exposure are measured. The test portfolios are the 10 long-run reversal portfolios ($N = 10$). The sample period is the one considered by LLM. (See the data section for a detailed description of the data and sample period.)

Panel A: Spurious Factor				
	LLMP		NP	
	0.950	0.900	0.950	0.900
$H = 1$	0.346	0.276	0.985	0.945
	[2.250]	[1.794]	[4.647]	[3.819]
$H = 4$	0.430	0.347	0.999	0.990
	[1.459]	[1.181]	[2.960]	[2.445]
$H = 8$	0.404	0.327	1.000	0.999
	[0.927]	[0.754]	[1.995]	[1.654]
Panel B: Capital Share Factor				
	LLMP		NP	
	0.950	0.900	0.950	0.900
$H = 1$	0.346	0.274	0.985	0.950
	[2.248]	[1.788]	[4.636]	[3.811]
$H = 4$	0.426	0.348	0.998	0.990
	[1.473]	[1.195]	[2.969]	[2.456]
$H = 8$	0.532	0.465	0.996	0.981
	[0.927]	[0.756]	[1.999]	[1.656]

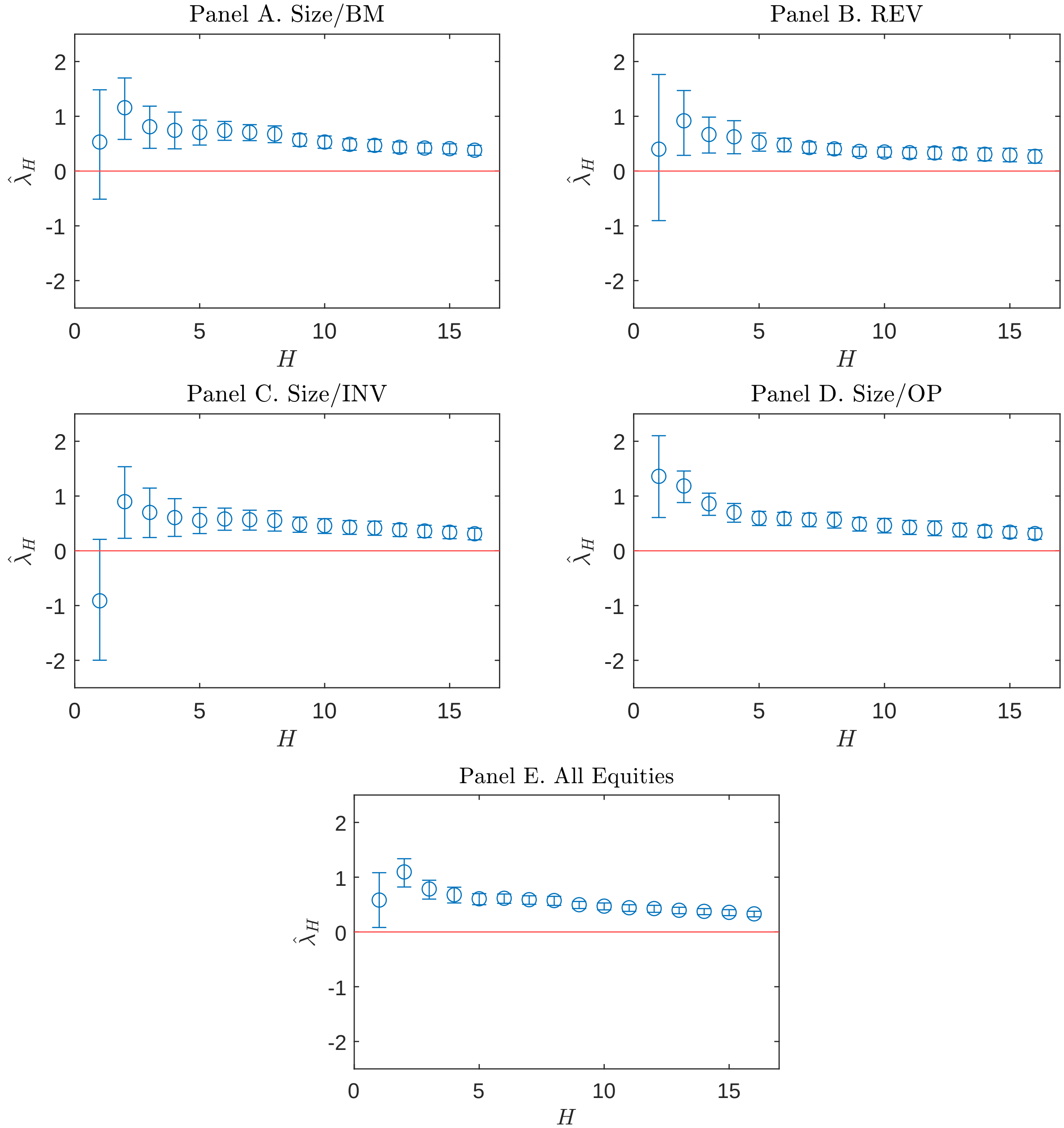

Figure 1. Expected return-beta regressions (LLM's bootstrap). The plots display LLM's capital share risk premium estimate, $\hat{\lambda}_H$, and its 95% bootstrap confidence interval for $H = 1, \ldots, 16$. H indicates the horizon in quarters over which returns and capital share exposure are measured. The number of bootstrap replications is 10,000. The sample period is the one considered by LLM. (See the data section for a detailed description of the data and sample period.)

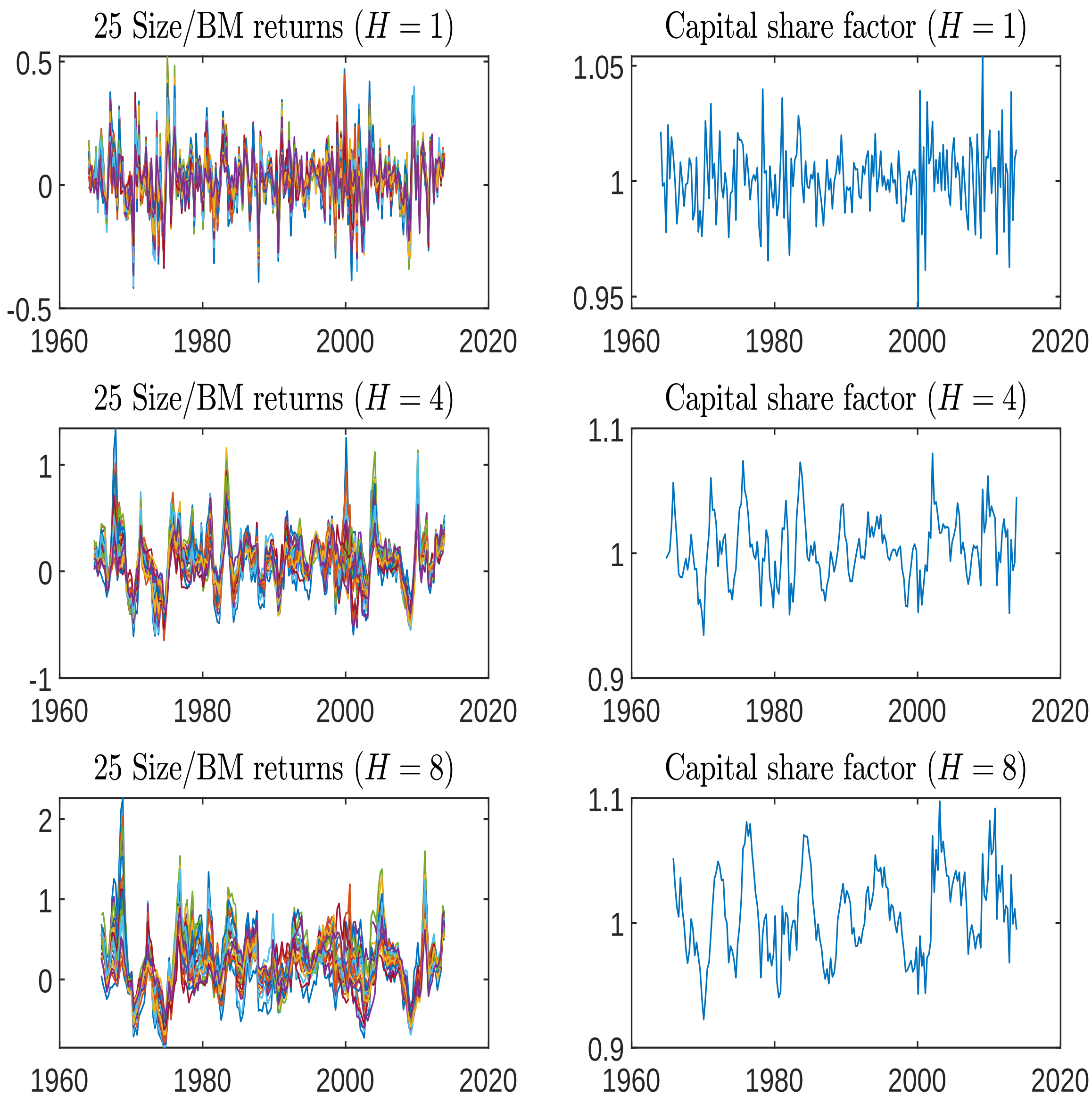

Figure 2. Time series dynamics of H-period excess returns (on 25 Size/BM portfolios) and capital share factor. The sample period for each H corresponds to the one considered by LLM. (See the data section for a detailed description of the data and sample period.)

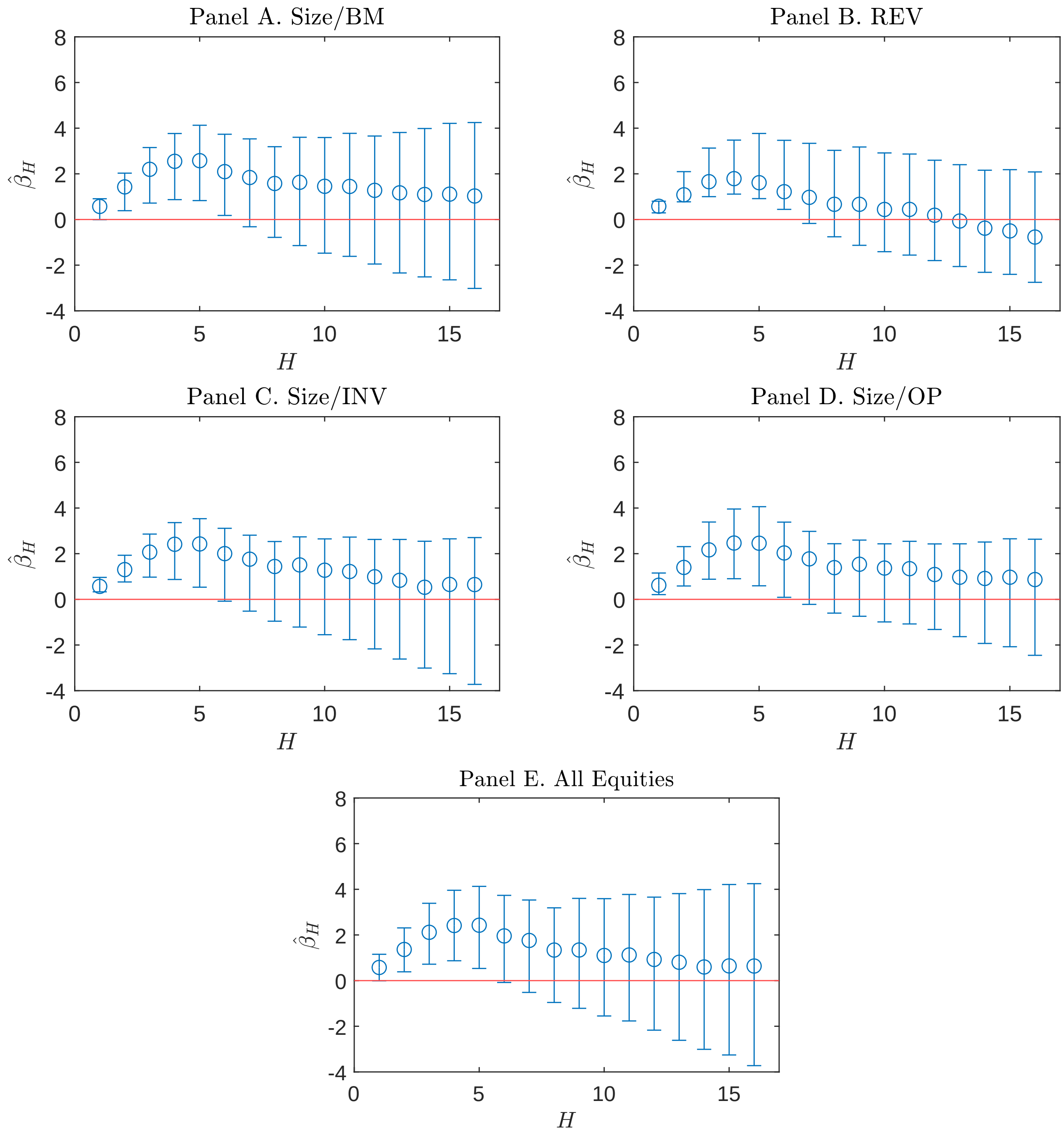

Figure 3. Factor Sensitivities. The plots display the median, min, and max values of the beta estimates, $\hat{\boldsymbol{\beta}}_H$ for $H = 1, \ldots, 16$. H indicates the horizon in quarters over which returns and capital share exposure are measured. The sample period is the one considered by LLM. (See the data section for a detailed description of the data and sample period.)

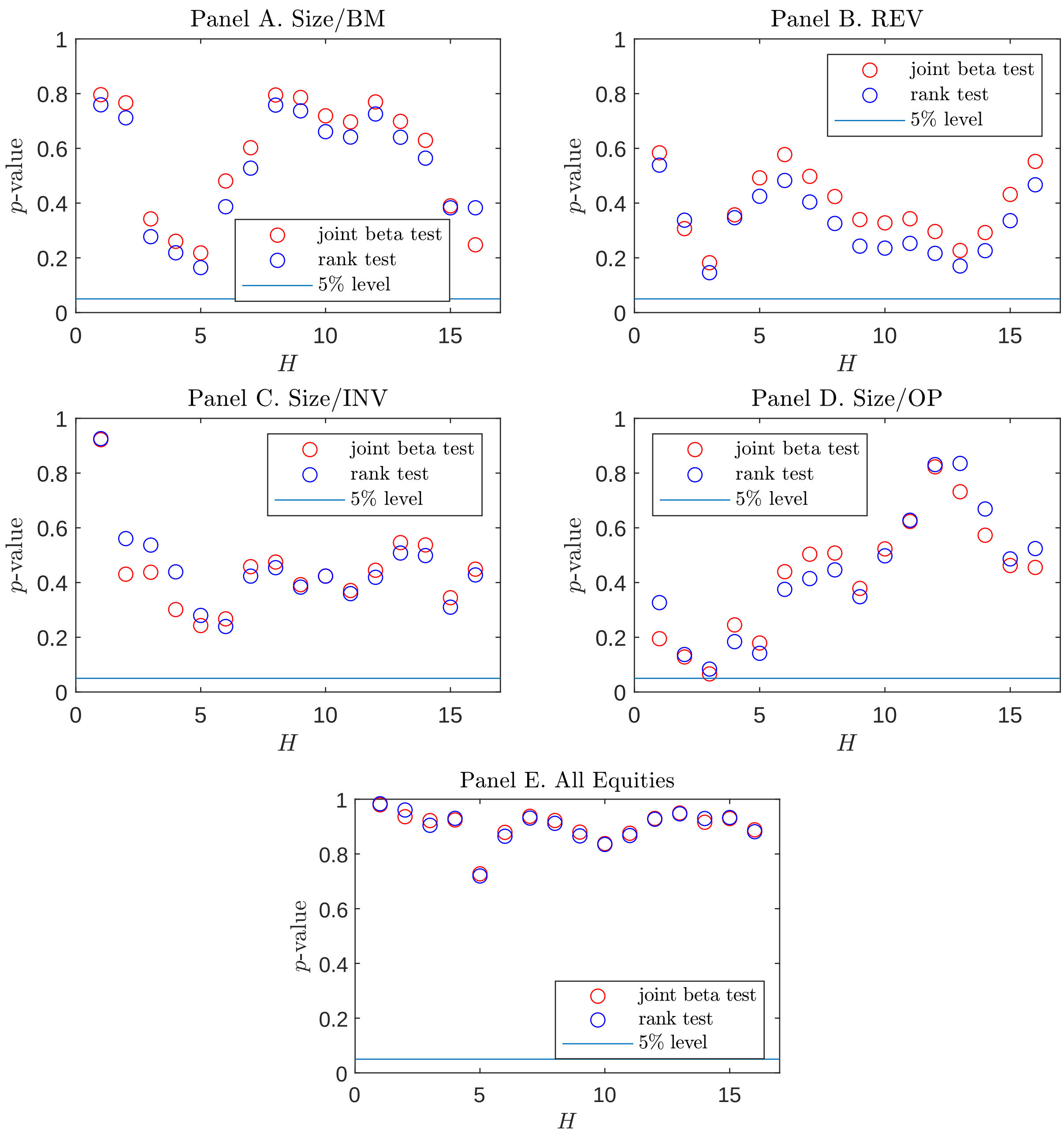

Figure 4. Joint beta and rank tests. The plots display the p-values of the bootstrap joint beta (red circles) and rank (blue circles) tests of model identification for $H = 1, \ldots, 16$. H indicates the horizon in quarters over which returns and capital share exposure are measured. The horizontal blue line represents the 5% nominal size of the tests. The number of bootstrap replications is 10,000. The sample period is the one considered by LLM. (See the data section for a detailed description of the data and sample period.)

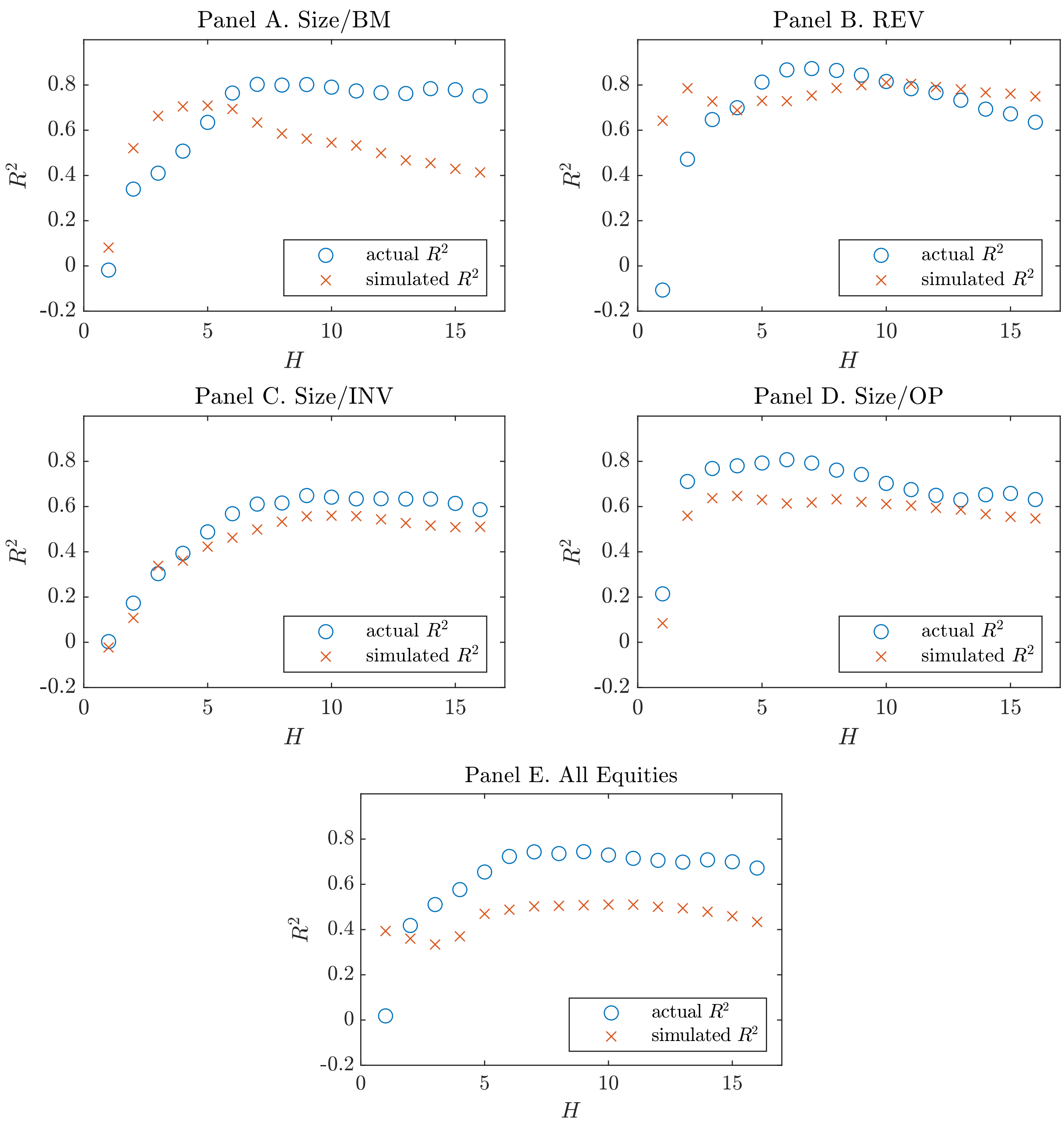

Figure 5. $\mathbf{R^2}$ profiles from actual and simulated data. The plots display the cross-sectional R^2 from the actual data and from simulated sample paths of a spurious factor for $H = 1, \ldots, 16$. H indicates the horizon in quarters over which returns and capital share exposure are measured. The number of simulated factor paths is 10,000. The test returns are kept fixed at their observed values across the simulations. The sample period is the one considered by LLM. (See the data section for a detailed description of the data and sample period.)

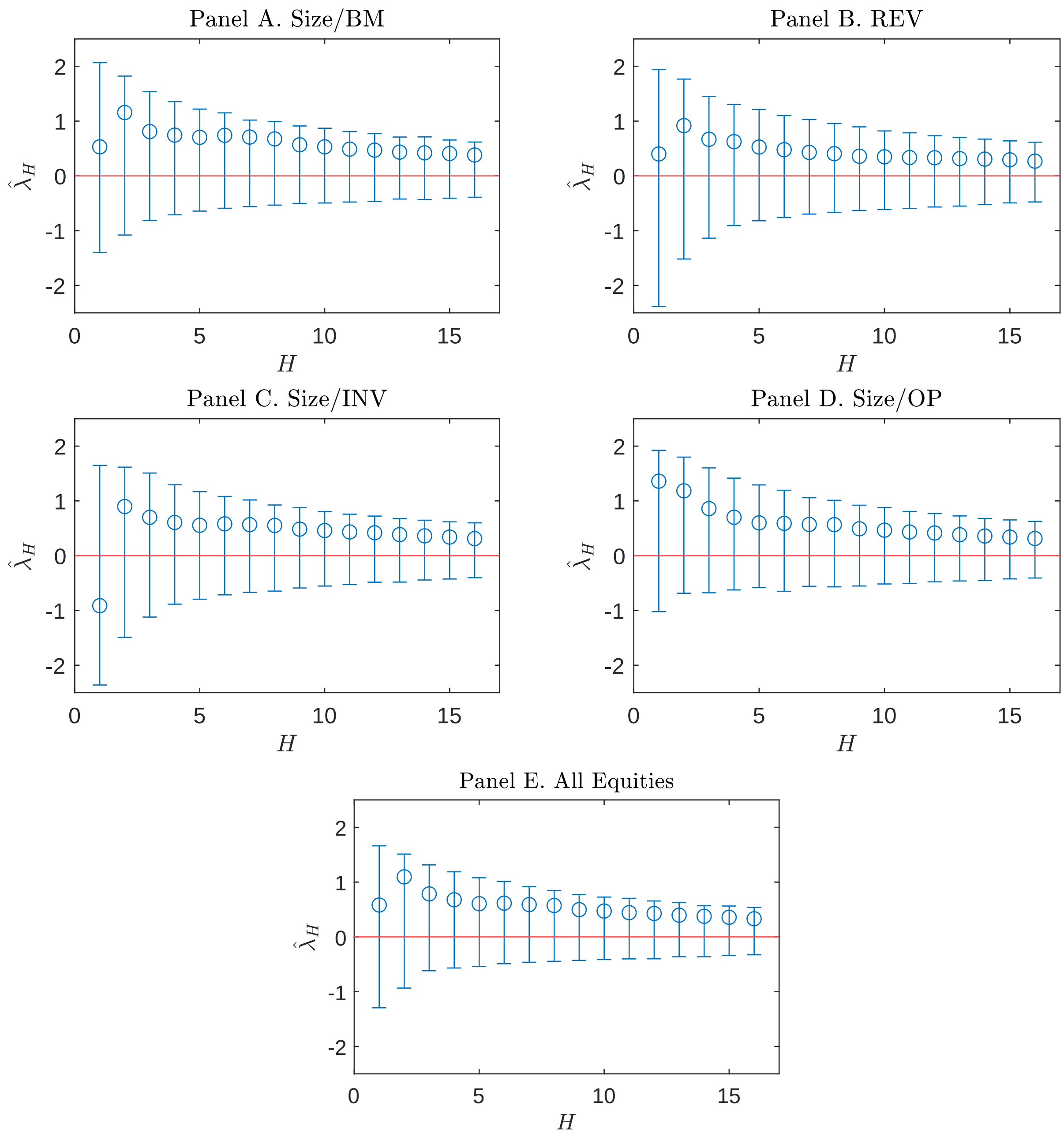

Figure 6. Expected return-beta regressions (nonparametric bootstrap). The plots display the capital share risk premium estimate, $\hat{\lambda}_H$, and its 95% (nonparametric) bootstrap confidence interval for $H = 1, \ldots, 16$. H indicates the horizon in quarters over which returns and capital share exposure are measured. The number of bootstrap replications is 10,000. The sample period is the one considered by LLM. (See the data section for a detailed description of the data and sample period.)