How easy is it for investment managers to deploy their talent in green and brown stocks?*

David Ardia[a], Keven Bluteau[b], Thien Duy Tran[a]

[a] *GERAD & Department of Decision Sciences, HEC Montréal, Montréal, Canada*
[b] *Department of Finance, Université de Sherbrooke, Canada*

Abstract

We explore the realized alpha-performance heterogeneity in green and brown stocks' universes using the peer performance ratios of Ardia and Boudt (2018). Focusing on S&P 500 index firms over 2014–2020 and defining peer groups in terms of firms' greenhouse gas emission levels, we find that, on average, about 20% of the stocks differentiate themselves from their peers in terms of future performance. We see a much higher time-variation in this opportunity set within brown stocks. Furthermore, the performance heterogeneity has decreased over time, especially for green stocks, implying that it is now more difficult for investment managers to deploy their skills when choosing among low-GHG intensity stocks.

Keywords: Greenhouse gas emissions (GHG), climate finance, carbon finance, peer performance

*We are grateful to IVADO and the Swiss National Science Foundation (grant #179281) for their financial support.

Email addresses: `david.ardia@hec.ca` (David Ardia), `keven.bluteau@usherbrooke.ca` (Keven Bluteau), `thien-duy.tran@hec.ca` (Thien Duy Tran)

1. Introduction

Environmental, social, and governance (ESG)-focused investing has become a hot topic in recent years. According to Morningstar, global sustainable fund assets hit a record US$3.9 trillion in 2021-Q3.[1] Among the ESG dimensions, the environmental concern is playing a leading role. As indicated by Pástor et al. (2021), 88% of the clients of BlackRock, the world's largest asset manager, rank environment as "the priority most in focus" among ESG criteria. Institutional investors and asset managers are increasingly tracking the greenhouse gas emissions of listed firms when building their portfolios (Krueger et al., 2020; Bolton and Kacperczyk, 2021).

In light of this, investment managers are increasingly subject to constraints regarding which assets they can invest in. Our research note aims to investigate to what extent they can differentiate themselves in terms of future performance when focusing on green or brown stocks' universes. We posit that the more heterogeneous a universe is, in terms of its underlying stocks' performance, the more easily good (bad) managers can deploy their skills (unskills) and differentiate themselves from their peers.

To explore the performance heterogeneity in green and brown stocks' universes, we rely on the peer performance ratios by Ardia and Boudt (2018). The output of their methodology is a triplet: (i) an equal-performance ratio, (ii) an underperformance ratio, and (iii) an outperformance ratio. The former measures the percentage of stocks in a given peer group that cannot be differentiated in terms of performance. The two other ratios attribute the remaining portion to over- or underperforming stocks. The approach is an extension to the false discovery rate approach by Storey (2002), pioneered by Barras et al. (2010) in their analysis of mutual funds performance; see Ardia and Boudt (2018) for an application to hedge funds. We refer to Harvey et al. (2020) for a review on multiple testing methods that have been developed to control for luck in active management.

Our empirical study focuses on S&P 500 index firms for 2014–2020. We use firms' greenhouse gas (GHG) emission intensity to create peer groups of green and brown stocks.[2] Using Carhart (1997) and Fama and French (2015) factor models, we find that, on average, about 20% of the stocks differentiate themselves from their peers in terms of realized-alpha performance over various horizons (from three months to one year of daily data). We see a much higher variability in this opportunity set within brown stocks. Furthermore, this heterogeneity has decreased over time, especially for green stocks, implying that it is now more difficult for investment managers to deploy their skills when allocating among low-GHG intensity stocks.

[1] See `https://www.reuters.com/business/sustainable-business/global-sustainable-fund-assets-hit-record`

[2] Gibson et al. (2021) find that the correlation between seven ESG-data providers is the highest for the environmental score (average correlation at 0.47), and rationalize this by the fact that greenhouse gas emissions are a key dimension of a firm's environmental performance.

2. Data

Our study focuses on firms in the S&P 500 index from January 2014 to December 2020. We define green (brown) firms as firms that create economic value while minimizing (not minimizing) damages that contribute to climate change. We follow Ardia et al. (2020) and use CO2-equivalent emissions scaled by the firm's revenues, namely, GHG emission intensity. We define total GHG emissions as the sum of scope 1 (direct CO2 emissions), scope 2 (indirect CO2 emissions), and scope 3 (indirect CO2 emissions from the value chain) in concordance with the Greenhouse Gas Protocol Standards.[3] CO2-equivalent emissions data come from Thomson/Refinitiv, and firms' revenues are from Compustat. Thus, our key ranking variable is the total tons of CO2-equivalent GHG emissions attributed to a one million dollar revenue, which we will refer simply to as GHG or GHG emissions intensity.

A summary of our data set is reported in Table 1. Since 2014, more than 60% of the S&P 500 firms have disclosed their CO2 emissions data. The distribution of GHG emissions intensity is highly positively skewed and fat-tailed.

[Insert Table 1 about here.]

Finally, daily stock prices data come from CRSP, and daily factor data are retrieved from Kenneth French's website.[4]

3. Methodology

To determine to what extent investment managers could differentiate themselves when investing in green or brown stocks in 2014–2020, we proceed as follows:

Step 1. In a given month, we form peer groups of brown and green stocks using the information available up to that month. We rely on the firms' latest GHG emissions released to form the groups.[5] Brown (green) stocks belong to the top 75th (bottom 25th) percentile of GHG intensities.[6] Each group contains the same number of peers.

Step 2. For each firm $i = 1, \ldots, N$ within a peer group, we compute the three peer-performance ratios, π_i^0, π_i^-, π_i^+, following Ardia and Boudt (2018). Expanding on Barras et al. (2010), Ardia and Boudt (2018) develop a triple-layered peer performance evaluation framework. For them, a stock can exhibit three types of peer performance with respect to a peer group:[7]

(i) equal-performance (π_i^0): the percentage of peers stocks that perform equally as stock i;

[3] See `https://ghgprotocol.org/standards`.

[4] See `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[5] GHG data are usually released with a one-year delay.

[6] We tested alternative thresholds, and results were qualitatively similar.

[7] Ardia and Boudt (2018) apply the methodology to hedge funds, but the approach can be applied to any set of instruments, such as stock returns as we do here.

(ii) underperformance (π_i^-): the percentage of peer stocks that outperform stock i;

(iii) outperformance (π_i^+): the percentage of peer stocks that underperform stock i.

For a given stock i, the peer performance ratios are obtained with a two-step procedure (see Ardia and Boudt, 2018, Section 2):

Step 2a. For each of the $N-1$ peer stocks $\mathcal{J}_i = \{j = 1, \ldots, N; j \neq i\}$, we calculate the p-value of the null hypothesis of equal-performance over a forward-looking evaluation period between stock i and stock $j \in \mathcal{J}_i$ using a pairwise test. We consider a h-month-ahead horizon of T daily returns and estimate the alpha differential from a factor model:

$$r_{i,t} - r_{j,t} = \Delta\alpha_{i,j} + \sum_{k=1}^{K} \beta_{i,j,k} F_{k,t} + \varepsilon_{i,j,t} \; (j \in \mathcal{J}_i \,;\, t = 1, \ldots, T)\,, \tag{1}$$

where $r_{i,t}$ is the daily return of stock i at time t, $F_{k,t}$ is the daily return of the kth factor at time t, and $\varepsilon_{i,j,t}$ is an error term. In our empirical application, we use the four-factor model by Carhart (1997) and the five-factor model by Fama and French (2015). For a given factor model, this leads to a set of coefficients $\{\Delta\hat{\alpha}_{i,j}, j \in \mathcal{J}_i\}$. Each coefficient is then standardized using the heteroscedastic and autocorrelation robust (HAC) standard error estimator of Andrews (1991) and Andrews and Monahan (1992), leading to a studentized test statistic from which the p-value is obtained via the probability integral transform.

Step 2b. The distribution of the p-values obtained in *Step 2a* is (asymptotically) a mixture of p-values that are uniformly distributed when the null hypothesis is true and p-values close to zero when the null hypothesis is false.[8] Following Ardia and Boudt (2018, Section 2.4), this allows us to set a data-driven cut-off point to estimate the proportion of equal-performance π_i^0 which is robust to false discoveries. The remaining proportion is attributed to the outperformance and underperformance ratios (Ardia and Boudt, 2018, Section 2.5).[9]

Step 3. Once the ratios are computed for the N firms within a peer group, we compute the aggregate ratios. First, $\pi^0 = \frac{1}{N}\sum_{i=1}^{N} \pi_i^0$ measures the equal-performance ratio within a peer group. As such, $1-\pi^0$ can be used to assess the extent to which stocks out- or underperform compared to their peers in the group: it is a measure of performance heterogeneity. The higher it is, the more direct it is for investment managers to deploy their talent at either (i)

[8] Barras et al. (2020) and Andrikogiannopoulou and Papakonstantinou (2019) recommend careful evaluation of the underlying data generating process when using the false discovery rate approach with financial data, especially when the sample size is small. This is not the case here, as we are dealing with a minimum of 60 observations (*i.e.*, three months of daily returns). Moreover, as noted in Ardia and Boudt (2018, Footnote 8), a very high correlation between the test statistics may lead to an inconsistent estimator. We checked this in our sample and found correlations in reasonable ranges (*i.e.*, within [-0.5,0.5] in the vast majority of the test statistics).

[9] All of the computations employed the R statistical computing language (R Core Team, 2020) with the package **PeerPerformance** (Ardia and Boudt, 2020), which is freely available at `https://CRAN.R-project.org/package=PeerPerformance`.

picking outperforming stocks or (ii) selling (or shorting) underperforming stocks in the peer group of brown or green funds to differentiate themselves from their competitors. Second, the quantities $\pi^- = \frac{1}{N}\sum_{i=1}^{N}\pi_i^-$ and $\pi^+ = \frac{1}{N}\sum_{i=1}^{N}\pi_i^+$ (which sum to $(1-\pi^0)$ by construction) are then used to assess the part of underperforming or outperforming stocks in the peer group.

Steps 1–3 are repeated every month starting from January 2014, yielding a time series of peer performance ratios. The ending date depends on the horizon h considered: September 2020 for $h = 3$, June 2020 for $h = 6$, and December 2019 for $h = 12$ months. Note that, by construction, the monthly ratios are autocorrelated, as they are based on overlapping daily returns (*e.g.*, in the $h = 12$ months horizon, 11 months of daily returns are overlapping between two consecutive months in the time series of the peer performance ratios).

4. Results

Following the steps in Section 3, we estimate the performance heterogeneity and the under- and outperformance ratios for the universes of green and brown stocks monthly over 2014–2020. Results are reported in Table 2.

First, we find that the unconditional performance heterogeneity is around 20%. The percentage is systematically higher for brown stocks than for green stocks. Results are robust over the three evaluation horizons (Columns 2–3, 4–5, and 6–7) and the two factors models (Panels A and B). Second, we observe a much higher variability in performance heterogeneity for brown stocks as reported by the standard deviation and the maximum-minimum range. For instance, at the three-month horizon with the four-factor model, the maximum performance heterogeneity percentage in the brown stocks group is 48.9%, while 29.8% for green stocks. This large difference is also observed with the five-factor model. Hence, the range of opportunity has been slightly higher for the brown stocks in 2014–2020 and has exhibited periods of higher heterogeneity for the set of brown stocks than for green stocks. In terms of investment opportunities, this means that it has been easier for investment managers to deploy their talent in brown stocks than green stocks. Finally, we see a negative trend in the performance heterogeneity (*i.e.*, the slope of a linear regression with time as the explanatory variable) for green and brown firms. The negative trend is significant at the one percent level for green stocks at the twelve-month horizon.[10] Results are robust to the inclusion of the performance heterogeneity of neutral firms (*i.e.*, firms with GHG intensity within the 25th–75th percentile range) as a control in the linear regression or when using a beta regression to measure the trend (Ferrari and Cribari-Neto, 2004).

When looking at the components of the performance heterogeneity, that is, the underperformance and outperformance ratios in Panels C and D, we see that it is primarily the underperformance ratios that drive the heterogeneity performance of brown stocks. For instance, in the case of the three-month horizon, the average is 11.6% for π^- and 9.4% for

[10] Statistical significance is based on heteroscedasticity and autocorrelation robust standard error estimators (Andrews, 1991; Andrews and Monahan, 1992).

π^+. It is also the case for green stocks, but to a lesser extent. We also see that the largest variability of both ratios is observed for the underperformance ratio. For both peer groups, the trend is negative but only significant for all trend specifications in the case of the outperformance ratio of green stocks over a twelve-month evaluation horizon. Overall, whether trading in the brown or green space, investment managers would have been able to differentiate themselves more easily by avoiding (or shorting) bad performer stocks. Moreover, picking outperforming green stocks has become more challenging in recent years.

[Insert Table 2 about here.]

In Figure 1, we display the time series of the performance heterogeneity obtained with the four-factor model setup over a three-month (top panels) or twelve-month (bottom panels) evaluation period.[11] We notice the declining percentage of performance heterogeneity for both brown and green stocks, as indicated by the trend in the dashed line. The higher variability is also evident for the brown stocks' universe. It is also interesting to note the dip in early 2020 for both universes in the three-month figure. The market crash in the first quarter of 2020 due to the COVID-19 pandemic drove the performance heterogeneity to its lowest value in green and brown universes. Post-crash, we see that the performance heterogeneity in green stocks remains particularly low compared to its historical average and brown stocks. Overall, while sustainable investing has become very trendy since the crisis, our findings suggest it is more challenging nowadays for investment managers in this space to differentiate themselves from their peers.

[Insert Figure 1 about here.]

[11] Results for the five-factor model are qualitatively similar.

References

Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858. doi:10.2307/2938229.

Andrews, D., Monahan, J., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 60, 953–966. doi:10.2307/2951574.

Andrikogiannopoulou, A., Papakonstantinou, F., 2019. Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power? Journal of Finance 74, 2667–2688. doi:10.1111/jofi.12784.

Ardia, D., Bluteau, K., Boudt, K., Inghelbrecht, K., 2020. Climate change concerns and the performance of green versus brown stocks. doi:10.2139/ssrn.3717722. Working paper.

Ardia, D., Boudt, K., 2018. The peer performance ratios of hedge funds. Journal of Banking & Finance 87, 351–368. doi:10.1016/j.jbankfin.2017.10.014.

Ardia, D., Boudt, K., 2020. **PeerPerformance**: Luck-corrected peer performance analysis in R. URL: https://github.com/ArdiaD/PeerPerformance.

Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. Journal of Finance 65, 179–216.

Barras, L., Scaillet, O., Wermers, R., 2020. Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power? A reply. Journal of Finance (Replication and Corrigenda), 1–34.

Bolton, P., Kacperczyk, M., 2021. Do investors care about carbon risk? Journal of Financial Economics 142, 517–549. doi:10.1016/j.jfineco.2021.05.008.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Financial Economics 52, 57–82. doi:10.1111/j.1540-6261.1997.tb03808.x.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22. doi:10.1016/j.jfineco.2014.10.010.

Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. Journal of Applied Statistics 31, 799–815. doi:10.1080/0266476042000214501.

Gibson, R., Krueger, P., Schmidt, P.S., 2021. ESG rating disagreement and stock returns. Financial Analysts Journal 77, 104–127. doi:10.1080/0015198x.2021.1963186.

Harvey, C.R., Liu, Y., Saretto, A., 2020. An evaluation of alternative multiple testing methods for finance applications. Review of Asset Pricing Studies 10, 199–248. doi:10.1093/rapstu/raaa003.

Krueger, P., Sautner, Z., Starks, L.T., 2020. The importance of climate risks for institutional investors. Review of Financial Studies 33, 1067–1111. doi:10.1093/rfs/hhz137.

Pástor, Ľ., Stambaugh, R.F., Taylor, L.A., 2021. Dissecting green returns. doi:10.2139/ssrn.3864502. Working paper.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Storey, J.D., 2002. A direct approach to false discovery rates. Journal of the Royal of Statistical Society B 64, 479–498. doi:10.1111/1467-9868.00346.

Table 1: Summary statistics of GHG

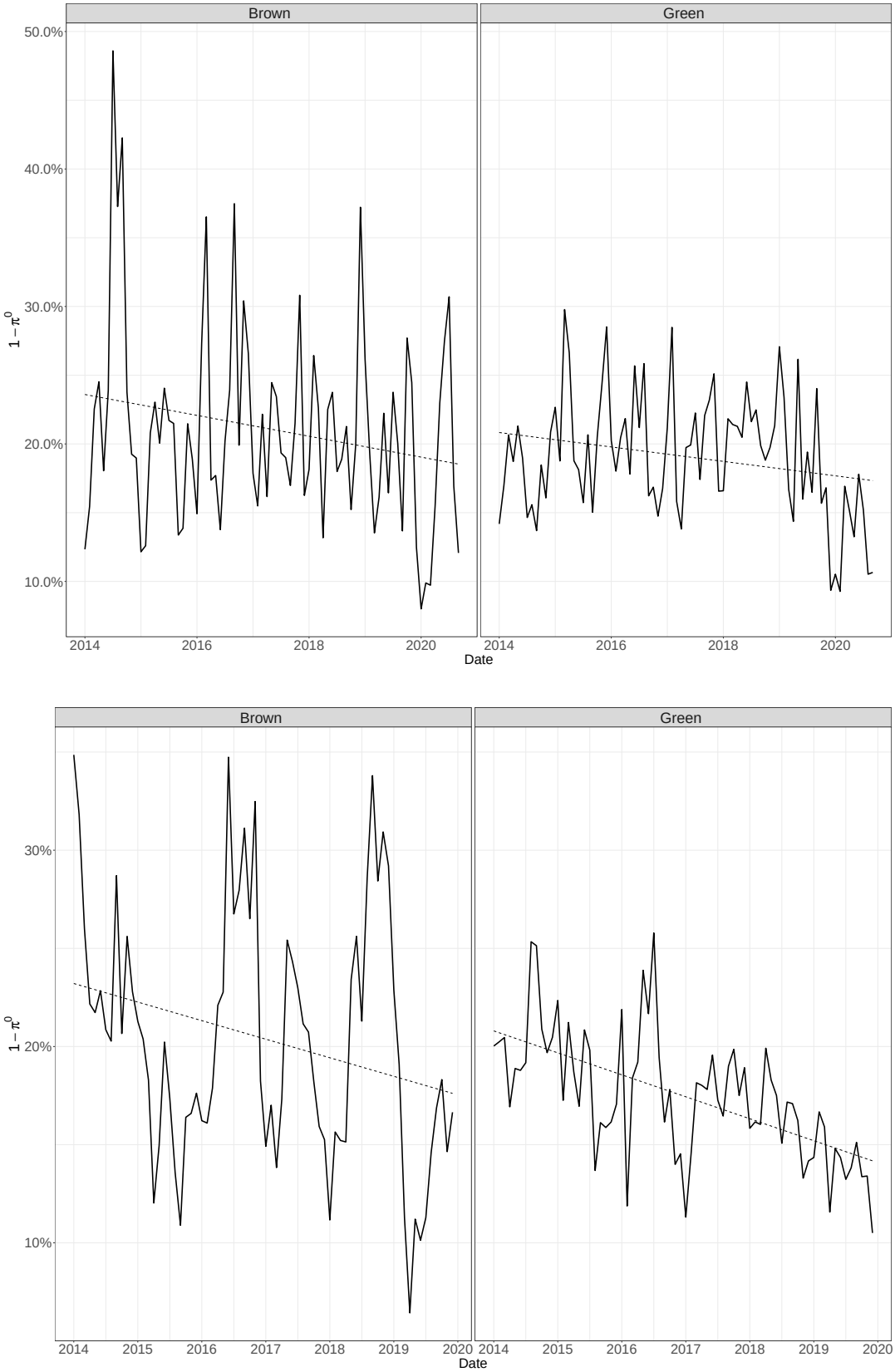
This table reports the summary statistics of the greenhouse gas emissions level scaled by firms' revenues (GHG emissions intensity). Total GHG emissions level is defined as the sum of scope 1 (direct CO2 emissions), scope 2 (indirect CO2 emissions), and scope 3 (indirect CO2 emissions from the value chain) in concordance with the Greenhouse Gas Protocol Standards. CO2 emissions are from Thomson/Refinitiv, and firms' revenues are from Compustat.

Panel A: Percentage of S&P 500 firms with CO2 emissions data	
Year	Value
2014	61.7
2015	62.2
2016	66.4
2017	70.2
2018	73.8
2019	80.2
2020	78.1
Panel B: Summary statistics of the GHG intensity data	
Statistics	Value
Average	811.6
Standard deviation	2,271.1
Skewness	4.4
Excess kurtosis	22.2
Minimum	0.0
25th percentile	37.2
50th percentile	77.1
75th percentile	358.6
Maximum	18,561.1

Table 2: Summary statistics of the peer performance ratios

This table reports the summary statistics of the performance heterogeneity $1-\pi^0$, the underperformance ratio π^-, and the outperformance ratio π^+. Panel A reports results for $1-\pi^0$ based on the alpha of the four-factor model by Carhart (1997). Panel B reports the results for the five-factor model by Fama and French (2015). Panel C (Panel D) reports the results for π^- (π^+) for the four-factor model. "Trend" reports the monthly trend of the ratio obtained by linear regression. "Trend with control" considers the performance heterogeneity of neutral funds as a control variable in the regression. "Trend with control (beta)" is a beta regression with the performance heterogeneity of neutral funds as a control variable in the regression. All trend coefficients are multiplied by 1,000 for readability. The symbols ***, ** and * indicate statistical significance of the trend at the 1%, 5%, and 10% levels, respectively, based on heteroscedasticity and autocorrelation robust standard error estimators (Andrews, 1991; Andrews and Monahan, 1992). Three-month ahead refers to peer performance ratios computed with the three-month daily returns following the evaluation month. Six-month ahead and one-year ahead refer to six months and twelve months of daily returns, respectively. Brown (Green) refers to the peer group of firms in the top 75th percentile (bottom 25th percentile) of the GHG emissions at the evaluation date. Setting with $h = 3$ months spans January 2014 to September 2020 (81 evaluation dates), $h = 6$ months spans January 2014 to June 2020 (79 evaluation dates), and $h = 12$ months spans January 2014 to December 2019 (72 evaluation dates).

Panel A: Four-factor model used to compute $1-\pi^0$						
	$h = 3$ months		$h = 6$ months		$h = 12$ months	
	Brown	Green	Brown	Green	Brown	Green
Average	21.1	19.1	20.0	18.1	20.4	17.5
Standard deviation	7.4	4.4	7.1	3.7	6.5	3.3
Minimum	8.0	9.3	6.4	10.0	6.4	10.5
Maximum	48.6	29.8	40.2	32.3	34.9	25.8
Trend	-0.6	-0.4	-1.2*	-0.4	-0.8	-0.9***
Trend with control	-0.2	-0.2	-0.6	-0.2	0.3	-0.6***
Trend with control (beta regression)	-1.1	-1.4	-4.0	-1.6	0.7	-4.6***
Panel B: Five-factor model used to compute $1-\pi^0$						
	$h = 3$ months		$h = 6$ months		$h = 12$ months	
	Brown	Green	Brown	Green	Brown	Green
Average	20.5	18.7	18.9	17.4	17.3	16.8
Standard deviation	7.8	4.7	6.9	4.2	7.3	4.2
Minimum	3.5	7.2	5.1	4.9	4.3	7.4
Maximum	46.5	32.9	38.9	28.0	42.6	27.5
Trend	-0.7	-0.5	-1.2*	-0.4	-1.3	-1.3***
Trend with control	-0.3	-0.2	-0.6	0.0	0.4	-0.7**
Trend with control (beta regression)	-2.2	-1.8	-4.0	-0.3	2.4	-5.5**
Panel C: Four-factor model used to compute π^-						
	$h = 3$ months		$h = 6$ months		$h = 12$ months	
	Brown	Green	Brown	Green	Brown	Green
Average	11.6	9.6	11.0	9.4	11.4	9.2
Standard deviation	5.3	3.3	5.0	2.7	5.1	2.4
Minimum	1.0	4.1	1.5	3.9	3.0	3.9
Maximum	28.0	18.9	25.8	18.2	22.9	14.8
Trend	-0.3	-0.2	-0.7	-0.2	-0.6	-0.4**
Trend with control	-0.1	-0.1	-0.5	-0.1	-0.1	-0.2
Trend with control (beta regression)	-1.3	-1.6	-5.3	-1.1	-0.5	-2.2
Panel D: Four-factor model used to compute π^+						
	$h = 3$ months		$h = 6$ months		$h = 12$ months	
	Brown	Green	Brown	Green	Brown	Green
Average	9.4	9.5	9.0	8.7	9.0	8.3
Standard deviation	3.6	2.7	3.3	2.6	3.0	2.2
Minimum	3.6	4.1	4.2	2.7	3.0	3.8
Maximum	20.6	15.7	16.3	15.8	16.5	13.2
Trend	-0.3	-0.2	-0.5***	-0.3	-0.2	-0.5***
Trend with control	-0.2	-0.1	-0.3*	-0.2	-0.1	-0.5**
Trend with control (beta regression)	-1.9	-1.7	-3.8	-2.0	-1.8	-5.7**

Figure 1: Performance heterogeneity over time

The following graphs present the evolution of performance heterogeneity $1 - \pi^0$ over time for the brown stocks' universe (left panel) and green stocks' universe (right panel). The performance heterogeneity is computed with three-month daily returns (top plots) and with twelve-month daily returns (bottom plots). Both are obtained with the four-factor model by Carhart (1997). The dashed line displays the time trend.