SN Bus Econ (2023) 3:176
https://doi.org/10.1007/s43546-023-00551-z

ORIGINAL ARTICLE

Does prior stock return correlation predict future stock return correlation?

Jonathan Ross[1]

Received: 10 June 2022 / Accepted: 8 August 2023 / Published online: 24 August 2023

Abstract

Using the sample time period 1950–2020, the historical average stock return correlation of a portfolio of firms is a strong positive predictor of that same portfolio's future average stock return correlation. The prediction improves when more prior monthly returns are used. The prediction holds after controlling for macroeconomic factors. The prediction also is much better for portfolios comprised of small firms relative to large firms. The prediction holds not only for the CRSP market of firms, but also for various-sized randomly chosen sub-samples of that market.

Keywords Stock return correlation · Predicting future returns · Return variances

JEL Classification G30 · G11

Introduction

Prior research has shown that beta varies over time despite the assumption of the Capital Asset Pricing Model that beta be constant.[1] However, betas for individual stocks, and more so for portfolios of stocks, do not vary much over time; that is beta is relatively persistent.[2] Therefore, knowing a portfolio's historical beta can help predict its future beta. Thus, one can identify a low-beta portfolio whose return has lower co-movement with the market return. However, if the stocks in this low-beta portfolio have higher comovement with each other, the investor is still susceptible to the risk that a large number of stocks in the chosen portfolio will drop in value in a given period, ceteris paribus. Therefore, the average pair-wise correlation in returns

[1] See Levy (1971), Roenfeldt et al. (1978), Theobald (1981) and Engle (2016), for example.
[2] See Kim (1993), DeJong and Collins (1985) and Chaudhuri and Lo (2015), for example.

✉ Jonathan Ross
jonathan.ross@wku.edu

[1] Gordon Ford College of Business, Western Kentucky University, 1906 College Heights Blvd, Bowling Green, KY 42101, USA

amongst the component stocks is of interest to a risk-averse investor and is a different metric than the portfolio's beta. Given two portfolios, A and B, A could have a lower beta than B but have a higher mean pair-wise correlation in returns than B. The adage "don't put all your eggs into one basket" applied to these two portfolios would seem to imply that B should be preferred to A. Choosing portfolio A is akin to putting more of your eggs into one basket since its component stocks are more correlated with each other on average than portfolio B.

A recent study presents an algorithm that identifies a portfolio of S stocks from the market portfolio each year whose mean pair-wise correlation (henceforth referred to as "average correlation") in historical returns[3] is minimized (Lee et al. 2021 chapter 48). The study shows that portfolios chosen with this 'correlation' algorithm are able to realize much smaller standard deviations than the market portfolio out-of-sample with insignificantly different returns, and the betas on such portfolios do not justify the returns they are able to earn. Their impetus for using the algorithm hinges on the assumption that portfolios with prior low average correlation will exhibit low future average correlation out-of-sample. This assumption is tested explicitly in this paper. Specifically, I am interested in the predictive ability of historical average return correlation for future average return correlation for any portfolio. Predicting future average correlation is an important endeavor for investors as portfolio standard deviation is increasing in average correlation (Lee et al. 2021, chapter 48 and Pollet and Wilson (2010), section 2.1). For a given level of return desired, risk-averse investors want to hold a portfolio which exhibits minimal portfolio standard deviation.[4] If average return correlation has predictive ability, then choosing a portfolio based on its realized prior average return correlation will give an investor insight into what that portfolio's future average return correlation will be, and hence what that portfolio's standard deviation (risk) will be out-of-sample.

I find that prior average return correlation is a statistically significantly positive predictor of future out-of-sample average return correlation for the CRSP market portfolio as well as for randomly drawn subsets of this market over a sample time period 1950-2020. Specifically, I regress the average monthly return correlation of the CRSP market in year t on the prior average return correlation using the prior 12, 24, 36, 48 or 60 monthly returns. This produces an average coefficient (t-stat) of 0.478(3.97) and an average R^2 of 15.48%. The coefficients, t-stats and R^2s improve when more prior returns were used. The results do not change qualitatively when the same regression is done with various-sized, randomly chosen subsets of the CRSP market ranging from 50 to 500 firms. I then repeat the regression on each of the CRSP market size quintiles and find an average coefficient (t-stat) of 0.706(6.88) and an average R^2 of 41.95% for the smallest size quintile. The same metrics for the largest size quintile are 0.22(1.63) and 2.82%. Therefore, there is a much stronger association between prior average return correlation and future average return correlation for groups of small firms relative to groups of large firms.

[3] Excess returns (returns less the risk-free rate).

[4] See for example (Markowitz 1952) and Clarke et al. (2006), among many others.

I next investigate which macroeconomic factors affect future average return correlation. The channel by which such factors would affect correlation necessarily needs to be a factor which captures how heterogeneous or homogeneuos changes in investor beliefs regarding the stock market are. If beliefs change homogeneously (everyone changes their belief in the same direction and approximately the same magnitude) or beliefs do not change, then correlation in average stock returns should increase, and vice versa if beliefs change heterogeneously. Also, such a factor must not be fully captured by prior average return correlation since that is already in the model. One such factor I find to be strongly negatively associated with future average return correlation is the lagged growth rate in U.S. household overall holdings of corporate equities.[5] When household corporate equity holdings increase in year *t*–1, the extra buying and selling of individual stocks that results over year *t* leads to lower correlation in average return correlation.[6]

The rest of this paper is organized as follows. Section 2 examines my hypotheses, section 3 explores the methodology I use to test the hypotheses, section 4 reports the results I find and section 5 concludes.

Hypotheses

Pollet and Wilson (2010) proposes that the average correlation between stocks may be a better predictor of the stock market risk premium than the stock market variance, because the former may better reveal changes in aggregate risk than the latter. Under certain assumptions and using the derivation in Campbell et al. (2002), chapter 2, Pollet and Wilson (2010) express the stock market risk premium as a function of both average correlation and average variance. They hypothesize that the coefficient on average correlation is statistically positive empirically. As evidence of this theoretically driven proposition, Pollet and Wilson (2010) find that the average correlation in prior daily returns for a portfolio of stocks is a statistically significantly positive predictor of next-quarter excess returns for that portfolio. Their study thus further highlights the importance of being able to predict average stock return correlation.

The findings from the prior research detailed above support the fact that average correlation in stock returns is an important empirical proxy for stock market risk. This motivates the endeavor of finding an empirical predictor of average correlation in stock returns. The first natural hypothesis is that historical average correlation in stock returns can predict such a proxy. I base this on the fact that this hypothesis held up in Lee et al. (2021), chapter 48, but was not directly tested.[7] I hypothesize, therefore, the following:

[5] Obtained from the Federal Reserve Bank of St. Louis website: https://fred.stlouisfed.org/.

[6] I examine other macroeconomic factors which could plausibly be thought to explain future average return correlation but find no statistically significant associations. The other factors I examine are the volatility index (VIX) created by the Chicago Board of Options Exchange, the economic policy uncertainty index (EPU) created by Baker et al. (2016) and the change in total assets of the federal reserve.

[7] The algorithm was designed to pick out groups of firms with low historical average correlation and these same groups subsequently realized low future average correlation out-of-sample.

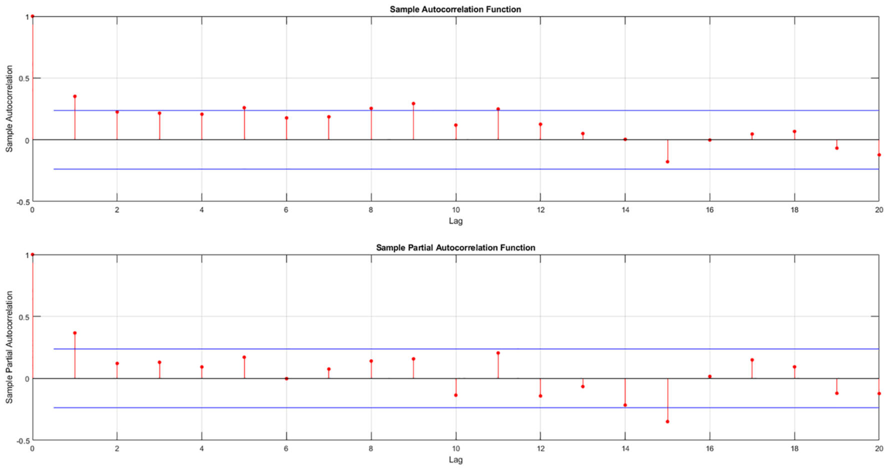

Fig. 1 Test of whether the correlation of $corr_t$ with various lags of itself is statistically different than zero. Horizontal bands are correlation levels deemed significant at a 5% level

H1a For a given group of firms, historical average correlation in excess stock returns is a statistically significantly positive predictor of future (out-of-sample) average correlation in excess stock returns.

Chan and Chen (1991) argue that the returns of small firms are more sensitive to macro-economic shocks and news in general compared to the returns of large firms. This is supported empirically by small firms having higher betas on average than large firms (see, for example, Ibbotson et al. (1997) Fig. 1 or Duffee (1995)). Beta is computed as the covariance of the firm's return with the market return divided by the variance of the market return as given in equation (1):

$$\begin{aligned} \beta_i &= \frac{COV_{i,M}}{\sigma_M^2} \\ &= \frac{\rho_{i,M} * \sigma_i * \sigma_M}{\sigma_M^2} \\ &= \rho_{i,M} \frac{\sigma_i}{\sigma_M} \end{aligned} \tag{1}$$

where $COV_{i,M}$ represents the covariance of firm i's return and the market return M over a given period, σ_M^2 is the variance of the market return and $\rho_{i,M}$ is the pearson correlation between firm i's return and the market return M. Empirically, it has been shown that $\rho_{i,M}$ is greater for small firms than for large firms and that σ_i is greater for small firms than for large firms. This is why prior research finds a higher expected return[8] for small firms than for large firms as these small firms have higher

[8] See Roll (1981), Reinganum (1982), Duffee (1995) and Ibbotson et al. (1997) for a few examples among many.

idiosyncratic risk (measured by the volatility of return, σ_i) and also higher systematic risk (measured by $\rho_{i,M}$). Since the returns of small firms are more correlated with the overall market return over a given period of time than the returns of large firms are, the returns of the small firms will move with each other over time more so than the returns of the large firms will move with each other. A useful analogy here is the following. Suppose you and I are each following a truck that is driving down the middle lane of a three-lane road who is constantly speeding up and slowing down; you are following in the left lane and I in the right lane. We each are trying to stay the same distance from the truck at all times. Since we are both following the truck and seeking to mimic its acceleration and deceleration pattern, our respective acceleration and deceleration patterns will be more correlated with each other than said patterns of two other cars who are also following the truck but not trying to remain the same distance from the truck at all times. The former two cars are the small firms and the latter two cars are the large firms and the truck is the market.

Ibbotson et al. (1997) and Mech (1993) offer a few other reasons why the returns of small firms will exhibit higher correlation over time with each other than the returns of large firms will. They argue that firm size most likely proxies for other factors that cause this effect. For example, size is related to the bid-ask spread; smaller firms have higher bid-ask spreads. The speed at which prices respond to new information is slower for high-spread(small) stocks than it is for low-spread (large) stocks because traders will not profit from acting upon the new information unless the value of the new information is higher than the costs of transacting.[9] Therefore, the returns of high-spread firms should move more together over time.

Since the returns on small-firm portfolios are more correlated (sensitive) with the market return (e.g. higher beta) than returns of large-firm portfolios, the average correlation amongst the small firms over time will be higher than the same metric for large firms. H1b formalizes my prediction regarding this correlation in average correlation over time.

H1b The effect posited in H1a is more pronounced for groups of smaller firms relative to groups of larger firms.

It is plausible that macroeconomic factors could affect future average return correlation and not be fully captured in prior average return correlation. The channel by which such factors would affect correlation necessarily needs to be a factor which captures how heterogeneous or homogeneous changes in investor beliefs regarding the stock market are. If beliefs change homogeneously (everyone changes their belief in the same direction and approximately the same magnitude) or beliefs do not change then average correlation in stock returns should increase and vice versa if beliefs change heterogeneously. For example, assume that no investor beliefs change. This means that whoever has shares of stock will hold them, and whoever does not will not seek to buy. Therefore, the prices of the stocks will all remain

[9] See Mech (1993) for a more complete explanation.

constant, and average correlation in returns tends towards 1 as the denominator (i.e. volatility) tends towards zero. Another scenario is when all beliefs change homogeneously such that, for example, mostly everyone believes the market is overvalued and seeks to sell. Every stock's price will then fall until equilibrium is reached and again, correlation will increase (since prices are all moving in the same direction) relative to a case when beliefs were more heterogeneous.

One such factor that captures changes in the beliefs of investors is the change in the total amount invested in corporate equities by households. If, in year $t-1$, total household holdings of corporate equities increase, and because of this there is more buying and selling in year t, average return correlation should decrease in year t as a result of the extra trading. To see why this should be expected, imagine that household holdings of corporate equities decreased to zero in year $t-1$. In year t, households do not hold any stocks and therefore there will be much less trading volume and the prices of the stocks will be less likely to move (only institutional investors hold them). This thereby increases the average correlation in year t relative to year $t-1$. Based on this reasoning, I hypothesize the following.

H2 Future average return correlation in year t is decreasing in the percentage change in household corporate equity holdings in year $t-1$.

Methodology

To test H1a, I first start with the CRSP universe over the time period 1945–2020 and eliminate any security whose share code isn't 10 or 11. I am only concerned with U.S.-based common stocks and do not want non-U.S.-based stocks or other securities such as real estate investment trusts or american depository receipts. This could cloud the analysis as they are thinly traded relative to the common stocks of U.S. companies. Most finance studies in general eliminate securities whose CRSP share codes are not 10 or 11 from their sample. I then estimate the following autoregressive (AR1) time-series regression over the sample period 1950–2020:

$$corr_t = \beta_0 + \beta_1 corr_{t-1} + \epsilon_t. \tag{2}$$

where $corr_{t-1}$ is the mean pair-wise correlation in historical monthly excess returns over the (12, 24, 36, 48 or 60) months prior to year t and $corr_t$ is the mean pair-wise correlation in the 12 monthly excess returns in year t. This specification requires each stock to have either 24, 36, 48, 60 or 72 consecutive monthly returns available in CRSP to be included. For example, if $corr_{t-1}$ is measured using the prior 12 monthly returns, then the regression observation for 1950 would include all stocks which have observations for each of the 24 consecutive months beginning January 1949 and ending December 1950. The mean pair-wise correlation of the portfolio over the 12 months in 1949 would be $corr_{t-1}$, while the mean pair-wise correlation for this same portfolio over the 12 months in 1950 would be $corr_t$. Because of this specification, there is a survivorship bias which is increasing in the number of

Table 1 Number of firms used in regression equs (2) and (3)

Years	12	24	36	48	60	Year	12	24	36	48	60
1950	936	908	871	814	775	1986	4365	3841	3133	2858	2470
1951	957	929	901	865	809	1987	4547	3990	3527	2885	2646
1952	981	953	925	897	861	1988	4866	4070	3583	3188	2610
1953	989	964	936	908	881	1989	5043	4409	3692	3266	2918
1954	997	979	954	926	899	1990	4915	4576	4023	3392	3007
1955	980	962	946	921	893	1991	4845	4514	4218	3718	3146
1956	972	958	940	924	900	1992	4710	4384	4115	3859	3423
1957	980	947	934	917	901	1993	4848	4374	4078	3835	3606
1958	973	950	918	905	888	1994	5078	4484	4056	3798	3582
1959	999	957	934	902	889	1995	5437	4654	4126	3745	3513
1960	992	971	932	910	880	1996	5634	4983	4280	3806	3471
1961	993	954	933	895	875	1997	5742	5073	4492	3865	3455
1962	1019	961	924	903	867	1998	5809	4993	4431	3964	3435
1963	1045	989	933	897	876	1999	5523	4969	4314	3847	3461
1964	1747	1013	958	905	869	2000	5198	4783	4323	3781	3390
1965	1780	1647	971	919	867	2001	4954	4507	4166	3799	3364
1966	1825	1683	1561	938	889	2002	4870	4447	4066	3769	3453
1967	1843	1708	1577	1464	891	2003	4532	4372	4024	3696	3445
1968	1811	1693	1564	1446	1345	2004	4326	4170	4023	3706	3414
1969	1835	1711	1599	1479	1366	2005	4128	3960	3823	3691	3413
1970	1921	1760	1641	1538	1425	2006	4089	3804	3649	3530	3405
1971	2067	1858	1700	1586	1488	2007	3957	3679	3426	3290	3183
1972	2155	1998	1797	1642	1535	2008	3866	3611	3370	3149	3033
1973	2229	2068	1920	1731	1581	2009	3816	3543	3318	3104	2909
1974	3566	2110	1958	1819	1648	2010	3633	3520	3278	3077	2888
1975	3548	3342	2030	1886	1758	2011	3462	3356	3257	3039	2855
1976	3222	3071	2896	1714	1604	2012	3356	3194	3105	3018	2826
1977	3220	2993	2860	2701	1621	2013	3266	3134	2986	2907	2825
1978	3455	2952	2752	2635	2495	2014	3230	3090	2966	2833	2759
1979	3437	3212	2760	2573	2471	2015	3208	3007	2880	2768	2648
1980	3389	3171	2974	2562	2396	2016	3235	2951	2770	2658	2560
1981	3376	3128	2929	2757	2385	2017	3197	3000	2737	2577	2475
1982	3503	3091	2868	2692	2543	2018	3157	3005	2819	2572	2424
1983	3838	3271	2897	2700	2538	2019	3140	2957	2816	2641	2418
1984	3882	3528	3015	2691	2512	2020	3185	2972	2796	2667	2502
1985	4326	3521	3210	2761	2468						
						Mean:	**3197**	**2905**	**2649**	**2423**	**2223**

This table summarizes the number of firms used in the calculation of each yearly regression observation over the (1950–2020) sample time period when (12, 24, 36, 48, 60) prior monthly excess returns are used in the equation (2) and (3) regression estimations months used for $corr_{t-1}$. For example, if I used the prior 60 months for $corr_{t-1}$ then the portfolio of stocks considered for the 1950 regression would be all those stocks with 72 consecutive monthly returns beginning January 1945 and ending December 1950. A stock that is missing any monthly return observation over this 6-year period would not be in the regression. However, the survivorship bias biases towards a portfolio of larger stocks. Newer stocks (less than 6 years old) are more likely to be small-cap relative to older stocks and these newer stocks will not be in the analysis when measuring $corr_{t-1}$ using the prior 60 monthly excess returns. This survivorship bias biases against finding the result hypothesized in H1b if H1a holds when I require a longer time series to compute $corr_{t-1}$. That is, the more prior returns I use in computing $corr_{t-1}$ the lower the number of small firms will be in the portfolio each year (since I classify firms into size quintiles as of 12/31 of year $t-1$), and thus the harder it will be to find evidence of H1b that $corr_t$ is increasing more in $corr_{t-1}$ for small firms than for large firms. Table 1 reports the number of firms used in the calculation of each yearly regression observation over the sample time period when 12, 24, 36, 48 or 60 monthly excess returns are used in Eq. (2).

To verify that an AR1 model is best for $corr_t$, I analyzed the autocorrelation and partial autocorrelation functions of $corr_t$ using Matlab's *autocorr* and *parcorr* functional tools. These functions compute the correlation of $corr_t$ with various lags of itself and these correlations are then tested for significance different from zero using an alpha significance level of 5%. The horizontal bands in each function's plot represent threshold correlation levels deemed significant at the 5% level. Notice in Fig. 1 that both the autocorrelation and partial autocorrelation functions spike when the lag is one period with subsequent lags being, for the most part, insignificant. Figure 1 provides evidence that modeling $corr_t$ as an AR1 process is best (Box et al. 2015).

Because equation (2) is an autoregresssive (AR1) model, it is important to test whether the series $corr_t$ follows a stationary process (e.g. has a mean and variance that does not depend on the time period over which it is measured). I used a standard (Dickey and Fuller 1979) test applied to the main regression equation (2) with all stocks included and over the entire sample time period to assess the stationarity of $corr_t$. The coefficient(t-stat) on $corr_{t-1}$ when regressing $\delta corr_t$ on $corr_{t-1}$ with an assumed constant was − 0.633 (− 5.49). The DF critical values for such a model at the 1% level of significance and a sample size of {50, 100} are {− 3.51,− 3.58}. Since my t-stat is less than both (my sample size is 1950 – 2020 = 70 years) I can safely reject the null hypothesis that $corr_t$ is a non-stationary time series in favor of the alternative that $corr_t$ is stationary.[10]

For H2, I use the lagged growth rate in U.S. household overall holdings of corporate equities. I obtain this from the Federal Reserve Bank of St. Louis website: https://fred.stlouisfed.org/. I estimate the following regression equation:

$$corr_t = \beta_0 + \beta_1 corr_{t-1} + \beta_2 \Delta hold_{t-1} + \epsilon_t. \tag{3}$$

[10] Whether I use 12,24,36,48 or 60 prior returns in measuring $corr_{t-1}$ the results of the DF test remain qualitatively the same.

Table 2 Prior average return correlation predicting future average return correlation (1950–2020)

Statistic	$m = 12$	$m = 24$	$m = 36$	$m = 48$	$m = 60$
$corr_t = \beta_0 + \beta_1 corr_{t-1} + \varepsilon_t$					
$\hat{\beta}_0$	0.119	0.118	0.093	0.082	0.074
	(5.061)***	(4.448)***	(3.326)***	(2.824)***	(2.631)***
$\hat{\beta}_1$	0.383	0.376	0.496	0.548	0.587
	(3.496)***	(2.819)***	(3.936)***	(4.415)***	(5.164)***
R^2	14.02%	10.84%	15.57%	17.46%	19.52%
F-stat	149.14	144.01	151.23	159.27	179.75

This table summarizes the results from estimating regression equation (2) in the paper over the sample period (1950–2020) where $corr_{t-1}$ is the mean pair-wise correlation in historical monthly excess returns over the $m = 12, 24, 36, 48, 60$ months prior to year t and $corr_t$ is the mean pair-wise correlation in the 12 monthly excess returns in year t. Excess returns are measured using the CRSP monthly returns less their respective monthly risk-free rates. At the beginning of each year, I start with the CRSP monthly returns universe and then keep only those securities with share codes 10 or 11 before estimating equation (2). Monthly risk-free rates are obtained from Ken French's website. Standard errors are robust to heteroskedasticity and t-statistics are reported in parentheses. ***(**)(*) indicate two-tailed statistical significance that the coefficient is not equal to zero at the 1%(5%)(10%) level, respectively

where $\Delta hold_{t-1}$ is the rate change in household total holdings in corporate equities from beginning of period $t-1$ to the end of period $t-1$. H2 predicts $\beta_2 < 0$.

Results

Table 2 reports the results from estimating the time-series regression equation (2). Notice the coefficient on $corr_{t-1}$ is highly significant regardless of how many prior months I use to estimate it. Specifically, $\beta_1 = 0.383$ when 12 prior monthly returns are used in calculating $corr_{t-1}$ and the t-stat is 3.496. β_1 increases to 0.587 when 60 prior monthly returns are used in calculating $corr_{t-1}$ with a t-stat of 5.164. The R^2's from estimating equation (2) increase from 14.02% when 12 prior monthly returns are used to 19.52% when 60 prior monthly returns are used. The evidence from Table 2 supports H1a that prior correlation in monthly excess returns is a strong predictor of 1-year ahead out-of-sample correlation in monthly excess returns for the CRSP market.

To assess the robustness of the results in Table 2, I next select various-sized random portfolios of stocks from the CRSP market and repeatedly estimate equation (2) for each portfolio over the sample time period. For each portfolio size, I only select100 random portfolios due to computing power. Specifically, I select 100 random portfolios of sizes $\{50, 100, 150, 200, 250, 300, 350, 400, 450, 500\}$ each and estimated equation (2) over the sample time period. For each given portfolio size, I average the respective statistics across the 100 random samples leaving me with 10 sets of portfolio mean statistics (e.g. mean β_0, and mean β_1 for the 100 random samples of portfolio size 50). I then compute the mean of these

Table 3 Prior average return correlation predicting future average return correlation – Randomly Chosen Firms (1950–2020)

	Statistic	$m = 12$	$m = 24$	$m = 36$	$m = 48$	$m = 60$
$corr_t = \beta_0 + \beta_1 corr_{t-1} + \varepsilon_t$ (100 random samples)						
Panel A	$\overline{\hat{\beta}_0}$	0.120	0.118	0.093	0.082	0.075
		(5.100)***	(4.439)***	(3.304)***	(2.817)***	(2.61)***
	$\overline{\hat{\beta}_1}$	0.379	0.376	0.496	0.548	0.587
		(3.472)***	(2.832)***	(3.906)***	(4.402)***	(5.082)***
	$\overline{R^2}$	13.80%	10.87%	15.44%	17.41%	19.35%
	$\overline{F-stat}$	146.00	140.41	147.54	156.2	175.09
$corr_t = \beta_0 + \beta_1 corr_{t-1} + \varepsilon_t$ (250 random samples)						
Panel B	$\overline{\hat{\beta}_0}$	0.120	0.118	0.093	0.082	0.075
		(5.174)***	(4.500)***	(3.361)***	(2.850)***	(2.647)***
	$\overline{\hat{\beta}_1}$	0.378	0.375	0.494	0.547	0.586
		(3.508)***	(2.859)***	(3.956)***	(4.435)***	(5.157)***
	$\overline{R^2}$	13.70%	10.79%	15.37%	17.28%	19.29%
	$\overline{F-stat}$	149.71	144.27	151.98	159.89	179.45

Panel A summarizes the averaged results from estimating regression equation (2) in the paper over the sample period (1950–2020) with 100 random samples each of sizes $N = 50, 100, 150, 200, 250, 300, 350, 400, 450, 500$ drawn from the CRSP market each year assuming a uniform distribution. I average both across the 100 random samples and across the various random sample sizes in reporting the statistics below (as indicated by the bars above them). Panel B repeats the analysis when I select 250 random samples. $corr_{t-1}$ is the mean pair-wise correlation in historical monthly excess returns over the $m = 12, 24, 36, 48, 60$ months prior to year t and $corr_t$ is the mean pair-wise correlation in the 12 monthly excess returns in year t. Excess returns are measured using the CRSP monthly returns less their respective monthly risk-free rates. At the beginning of each year, I start with the CRSP monthly returns universe then keep only those securities with share codes 10 or 11. From that population of firms I choose 100 random samples of the aforementioned sizes and calculate $corr_{t-1}$ and $corr_t$ using excess returns. Monthly risk-free rates are obtained from Ken French's website. Standard errors are robust to heteroskedasticity and average t-statistics are reported in parentheses. ***(**)(*) indicate two-tailed statistical significance that the coefficient is not equal to zero at the 1%(5%)(10%) level respectively

means (e.g. the mean β_0 of the 100 random samples of size 50, ..., 500). Table 3 Panel A reports the average results from this procedure. It is important to point out that the mean of a given regression statistic across the various portfolio sizes for a fixed number of random samples was basically constant. When I took 100 random samples of size 50, for example, the mean β_1 of those 100 random samples was the same as when the portfolio size was 100 or 150 or 500.

Notice from Table 3 Panel A that, on average across the various random sample sizes, the same results hold qualitatively as in Table 2 where the sample was the

Table 4 Prior average return correlation and future average return correlation by size (1950–2020)

	Statistic	$m = 12$	$m = 24$	$m = 36$	$m = 48$	$m = 60$
$corr_t = \beta_0 + \beta_1 corr_{t-1} + \varepsilon_t$						
Q1	$\widehat{\beta}_0$	0.052	0.048	0.033	0.017	0.022
		(3.759)***	(3.094)***	(2.070)**	(1.096)	(1.317)
	$\widehat{\beta}_1$	0.627	0.619	0.711	0.807	0.768
		(6.137)***	(5.589)***	(7.238)***	(8.266)***	(7.188)***
	R^2	40.64%	36.53%	40.71%	47.22%	44.65%
	$F-stat$	116.08	111.79	119.31	126.84	126.84
Q2	$\widehat{\beta}_0$	0.097	0.094	0.070	0.051	0.051
		(4.763)***	(3.866)***	(2.806)***	(2.039)**	(1.981)**
	$\widehat{\beta}_1$	0.461	0.472	0.600	0.694	0.701
		(4.185)***	(3.495)***	(4.931)***	(5.672)***	(6.052)***
	R^2	20.70%	17.53%	23.88%	29.32%	30.38%
	$F-stat$	131.81	130.33	139.70	147.73	163.06
Q3	$\widehat{\beta}_0$	0.144	0.150	0.115	0.101	0.086
		(5.098)***	(4.673)***	(3.291)***	(2.776)***	(2.432)***
	$\widehat{\beta}_1$	0.354	0.328	0.484	0.545	0.619
		(3.154)***	(2.358)**	(3.387)***	(3.832)***	(4.684)***
	R^2	11.54%	7.58%	13.05%	15.27%	19.14%
	$F-stat$	175.79	162.45	173.28	183.55	206.01
Q4	$\widehat{\beta}_0$	0.190	0.199	0.157	0.153	0.138
		(5.922)***	(5.281)***	(4.043)***	(3.665)***	(3.417)***
	$\widehat{\beta}_1$	0.227	0.196	0.364	0.385	0.447
		(1.922)**	(1.312)	(2.522)***	(2.624)***	(3.263)***
	R^2	4.72%	2.48%	6.69%	6.71%	8.34%
	$F-stat$	174.77	173.20	180.75	183.29	196.87
Q5	$\widehat{\beta}_0$	0.213	0.230	0.209	0.205	0.201
		(6.589)***	(5.943)***	(4.754)***	(4.127)***	(3.942)***
	$\widehat{\beta}_1$	0.216	0.153	0.231	0.246	0.265
		(2.088)**	(1.137)	(1.630)	(1.597)	(1.695)**
	R^2	4.49%	1.59%	2.75%	2.63%	2.66%
	$F-stat$	185.47	177.08	180.97	190.54	192.27

This table summarizes the results from estimating regression equation (2) in the paper for each size quintile over the sample period (1950–2020) where $corr_{t-1}$ is the mean pair-wise correlation in historical monthly excess returns over the $m = 12, 24, 36, 48, 60$ months prior to year t and $corr_t$ is the mean pair-wise correlation in the 12 monthly excess returns in year t. Excess returns are measured using the CRSP monthly returns less their respective monthly risk-free rates. At the beginning of each year, I start with the CRSP monthly returns universe and then keep only those securities with share codes 10 or 11. Next, I put the firms into size quintiles based on their respective market capitalizations as of 12/31 of the prior year before estimating equation (2). Q1 (Q5) represent the smallest(largest) size quintiles. Monthly risk-free rates are obtained from Ken French's website. Standard errors are robust to heteroskedasticity and t-statistics are reported in parentheses. ***(**)(*) indicate two-tailed statistical significance that the coefficient is not equal to zero at the 1%(5%)(10%) level respectively

Table 5 U.S. household corporate Equity holdings and future average return correlation (1950–2020)

Statistic	m = 12	m = 24	m = 36	m = 48	m = 60
$corr_t = \beta_0 + \beta_1 corr_{t-1} + \beta_2 \Delta hold_{t-1} + \varepsilon_t$					
$\hat{\beta}_0$	0.129	0.126	0.106	0.096	0.088
	(5.601)***	(4.994)***	(3.856)***	(3.216)***	(3.004)***
$\hat{\beta}_1$	0.410	0.417	0.497	0.536	0.583
	(3.864)***	(3.287)***	(4.175)***	(4.394)***	(5.040)***
$\hat{\beta}_2$	−0.163	−0.176	−0.143	−0.127	−0.134
	(−2.852)***	(−3.238)***	(−2.582)**	(−2.229)**	(−2.467)**
R^2	21.92%	19.73%	21.31%	21.84%	24.44%
$F - stat$	111.58	111.81	110.92	114.59	129.18

This table summarizes the results from estimating regression equation (3) in the paper over the sample period (1950–2020) where $corr_{t-1}$ is the mean pair-wise correlation in historical monthly excess returns over the $m = 12, 24, 36, 48, 60$ months prior to year t and $corr_t$ is the mean pair-wise correlation in the 12 monthly excess returns in year t. Excess returns are measured using the CRSP monthly returns less their respective monthly risk-free rates. At the beginning of each year I start with the CRSP monthly returns universe and then keep only those securities with share codes 10 or 11 before estimating equation (3). $\Delta hold_{t-1}$ is the rate change in U.S. household total holdings of corporate equities over year $t - 1$ defined as the beginning holdings less the ending holdings divided by the beginning holdings. I obtain U.S. household total holding in corporate equities data from the Federal Reserve Bank of St. Louis website: https://fred.stlouisfed.org/. Monthly risk-free rates are obtained from Ken French's website. Standard errors are robust to heteroskedasticity and t-statistics are reported in parentheses. ***(**)(*) indicate two-tailed statistical significance that the coefficient is not equal to zero at the 1%(5%)(10%) level respectively

whole CRSP market over the sample time period. As a robustness check, I repeat the procedure described above but took 250 random samples for each of the various portfolio sizes and averaged the results from these samples across the portfolio sizes.[11] These results are reported in Table 3 Panel B. Notice the results reported in Table 3 Panel B are virtually identical to those reported in Table 3 Panel A. The results summarized in Table 3 provide further evidence in favor of H1a that prior correlation in monthly excess returns is a strong predictor of one-year ahead out-of-sample correlation in monthly excess returns for any portfolio in general (not just the CRSP market portfolio).

Table 4 summarizes the results from estimating regression equation (2) by quintile over the sample time period. Notice that for each m, there is a monotonic decrease in β_1, the $t - stat$ for β_1 and the R^2 moving from size quintile $Q1$ to $Q5$. For example, when using the correlation in prior 12 monthly returns to predict the correlation in the future 12 monthly returns, β_1 falls from 0.627 to 0.216 moving from size quintile $Q1$ to $Q5$ and the $t - stat$ falls from 6.137 to 2.088. Moreover, the R^2 falls from 40.64% to 4.69%. When the correlation in 48 prior monthly returns are used to estimate the correlation in the future, 12 monthly returns, β_1 falls from

[11] I couldn't take more than 250 random samples and repeat the analysis as this exceeded my computing power.

0.807 to 0.246 from the smallest to largest quintile. The $t - stat$ falls from a strongly significant 8.266 to an insignificant 1.597 and the R^2 falls from 47.22% to 2.63%. In summary, Table 4 provides strong evidence in favor of H1b and I conclude that the ability of prior excess return correlation to predict future excess return correlation is much higher for smaller firms relative to larger firms.

Table 5 summarizes the results from estimating regression equation (3). Notice the coefficient on $corr_{t-1}$ remains highly significant after the addition of $\Delta hold_{t-1}$. The coefficient on $\Delta hold_{t-1}$ is highly significantly negative, confirming H2. Notice also that the adjusted R^2 increases around 5-8% relative to the Table 2 estimates of the equation (2) specification. This result holds regardless of how many prior months I use to estimate $corr_{t-1}$. The evidence from Table 5 supports H2 that the prior year growth rate in U.S. household holdings of corporate equities is negatively associated with 1-year ahead out-of-sample correlation in monthly excess returns for the CRSP market. The evidence from Table 5 also suggests that the equation (3) specification is a better model than the equation (2) specification.

I also examine other macroeconomic factors that plausibly could affect future average return correlation. For example, I test whether the volatility index (VIX) created by the Chicago Board of Options Exchange can explain future average return correlation. The idea is that VIX measures overall market volatility and possibly the risk perceived by investors. When this increases in year $t - 1$, I would expect an increase in average return correlation in year t as investors pull out of the market in year t uniformly (greater homogeneity of beliefs).However, I find no statistical relationship between VIX in the prior year and future average return correlation. One reason may may be the greatly reduced sample size. The VIX measure is only available since 1990 and thus the estimation of equation (3) with ΔVIX_{t-1} substituted for $\Delta hold_{t-1}$ has only 31 observations instead of the 71 that regression equation (3) has when using the household corporate equity holding data.

Similarly, I also examine whether the economic policy uncertainty (EPU) index created by Baker et al. (2016) can explain future average return correlation.[12] Increases in uncertainty surrounding economic policies will make individuals less likely to either buy or sell and thus correlation in future average returns would increase. However, I find no statistical relationship between EPU in the prior year and future average return correlation.

Finally, I also examine whether changes in the total assets held by the U.S. federal reserve can explain future average return correlation. Large changes in the total assets held by the FED proxies for unconvential monetary policies, as put forth in Cortes et al. (2022) and Dedola et al. (2021), which would be expected to make people more uncertain regarding the future. The effect is possibly that investors either stop buying and selling or they all seek to sell (uniformly). Either of these two would increase correlation in future average returns. However, again I find no statistical relationship between large changes in the total assets held by the FED in the prior year and future average return correlation. I conclude that either the effect of VIX, EPU and change in FED assets on future average return correlation is already

[12] Obtained from their website: https://www.policyuncertainty.com/.

captured by prior average return correlation, or I just do not have enough power in the regression to establish a statistical relationship due to a small sample size of 71 yearly regression observations.[13]

Conclusion

The average pair-wise correlation of a portfolio's component stock returns is an important metric for a risk-averse investor. In this paper, I find evidence that the average correlation in historical excess stock returns for a portfolio strongly predicts the future average correlation in excess stock returns for this same portfolio. Furthermore, the ability of prior stock return correlation to predict future stock return correlation is highest for portfolios comprised of small stocks relative to large stocks. Additionally, the growth rate in U.S. household holdings of corporate equities is a negative predictor of future average correlation. That is, future average correlation of the returns of the market portfolio of stocks declines when U.S. households increase their holdings of corporate equities. The strong positive association between prior average return correlation and future average return correlation remains after controlling for this macroeconomic factor. This finding is useful for a risk-averse investor who seeks to minimize the average pair-wise return correlation in their portfolio and shield themselves as much as possible from idiosyncratic risk while still earning a respectable return compensating them fairly for the systematic risk they assume when holding a stock portfolio.

Acknowledgements None.

Author contributions I am the sole author and the only one who contributed.

Funding None.

Data availibility The datasets generated during and/or analyzed during the current study are available in the CRSP monthly returns file which can be obtained with a subscription to Wharton Research Data Services from the University of Pennsylvania.

Declarations

Conflict of interest There is no conflict of interest between this article's content and any third party

Ethical approval This article does not contain any studies with human participants performed by the author.

References

Baker SR, Bloom N, Davis SJ (2016) Measuring economic policy uncertainty. Q J Econ 131(4):1593–1636

[13] The coefficients on each of those variables separately were close to being significant at the 10% level (untabulated analysis).

Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis: forecasting and control. John Wiley & Sons

Campbell JY, Viceira LM, Viceira LM et al. (2002). *Strategic asset allocation: portfolio choice for long-term investors*

Chan K, Chen N-F (1991) Structural and return characteristics of small and large firms. J Fin 46(4):1467–1484

Chaudhuri SE, Lo AW (2015). Spectral analysis of stock-return volatility, correlation, and beta. In *2015 IEEE Signal Processing and Signal Processing Education Workshop (SP/SPE)*, pp. 232–236

Clarke RG, De Silva H, Thorley S (2006) Minimum-variance portfolios in the us equity market. J Portf Manag 33(1):10–24

Cortes GS, Gao GP, Silva FB, Song Z (2022) Unconventional monetary policy and disaster risk: Evidence from the subprime and covid-19 crises. J Int Money Financ 122:102543

Dedola L, Georgiadis G, Gräb J, Mehl A (2021) Does a big bazooka matter? quantitative easing policies and exchange rates. J Monet Econ 117:489–506

DeJong DV, Collins DW (1985) Explanations for the instability of equity beta: Risk-free rate changes and leverage effects. J Financ Quant Anal 20(1):73–94

Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 74(366a):427–431

Duffee GR (1995) Stock returns and volatility a firm-level analysis. J Financ Econ 37(3):399–420

Engle RF (2016) Dynamic conditional beta. J Financ Economet 14(4):643–667

Ibbotson RG, Kaplan PD, Peterson JD (1997) Estimates of small-stock betas are much too low. J Portf Manag 23(4):104

Kim D (1993) The extent of nonstationarity of beta. Rev Quant Financ Acc 3(2):241–254

Lee C-F, Lee A, Lee J (2021). *Handbook of investment analysis, portfolio management, and financial derivatives*

Levy RA (1971) On the short-term stationarity of beta coefficients. Financ Anal J 27(6):55–62

Markowitz H (1952) Portfolio selection. J Financ 7:77–91

Mech TS (1993) Portfolio return autocorrelation. J Financ Econ 34(3):307–344

Pollet JM, Wilson M (2010) Average correlation and stock market returns. J Financ Econ 96(3):364–380

Reinganum MR (1982) A direct test of roll's conjecture on the firm size effect. J Financ 37(1):27–35

Roenfeldt RL, Griepentrog GL, Pflaum CC (1978) Further evidence on the stationarity of beta coefficients. J Financ Quantit Analy 13(1):117–121

Roll R (1981) A possible explanation of the small firm effect. J Financ 36(4):879–888

Theobald M (1981) Beta stationarity and estimation period: some analytical results. J Financ Quant Analy 16(5):747–757