

# THE ASHLEY AND PATTERSON (1986) TEST FOR SERIAL INDEPENDENCE IN DAILY STOCK RETURNS, REVISITED

RICHARD A. ASHLEY, FAEZEH NAJAFI, AND DOUGLAS M. PATTERSON

**ABSTRACT.** We update and extend the non-parametric test proposed in Ashley and Patterson (1986) – of the proposition that the (pre-whitened) daily stock returns for a firm are serially independent, and hence unforecastable from their own past. That paper applied this test to daily returns from 1962 to 1981 for several U.S. corporations and aggregate indexes, finding mixed evidence against this serial independence hypothesis. The returns dataset is updated here to include thirteen firms which are currently more relevant, and the sample is extended through the end of 2023. We also update the simulation methodology here to properly account for the conditional heteroscedasticity in the daily returns data, so that the present results should now be more statistically reliable. The results are broadly in line with our earlier results, but do provide some surprises, including a re-framing of the test; the new results also suggest further avenues of research in this area.

JEL Classification Codes: C18, C22, C31, C51, C52

Author contact information.<sup>123</sup>

## 1. Introduction

The present work updates, reapplies and – notably – re-frames the non-parametric test of the proposition that the (pre-whitened) daily returns from a share of corporate stock are serially independent, and hence not forecastable from their own past, which was proposed in Ashley and Patterson (1986). This earlier work will be referred to as AP (1986) below.

The daily returns for some stocks are, of course, well known to be serially correlated, albeit to only a very modest degree. The AP (1986) test was consequently applied to the residuals of an  $AR(p)$  ‘pre-whitening’ model fitted to the sample data, and the ‘forecastability’ proposition referenced above – related, of course, to the Efficient Markets Hypothesis – is

---

*Date:* June 27, 2024.

*Key words and phrases.* Stock Returns, Random Walks, Serial Independence, Bootstrap, Nonparametric Testing, Serial Independence.

<sup>1</sup>*Ashley:* Department of Economics, Virginia Tech, Blacksburg, VA, 24061. E-mail: ashleyr@vt.edu. (Corresponding Author)

<sup>2</sup>*Najafi:* Department of Economics, Virginia Tech, Blacksburg, VA, 24061. E-mail: fiza@vt.edu.

<sup>3</sup>*Patterson:* Department of Finance (Emeritus), Virginia Tech, Blacksburg, VA, 24061. E-mail: amex@vt.edu.

with respect to forecastability beyond the almost negligible amount provided by this linear  $AR(p)$  model.

The present paper proceeds as follows:

- In Section 2 we briefly recap the random walk model for stock prices and the AP (1986) non-parametric test of the serial independence null hypothesis with regard to pre-whitened daily returns for a particular stock time-series of length  $T$ .<sup>4</sup> This test was based on a bootstrap simulation of the distribution of the maximum absolute excursion such a random walk will exhibit after  $T$  serially independent steps.

Section 2 briefly reviews the AP (1986) random walk test, primarily via an illustrative graphical example using recent daily returns data on the 3M Company. We integrate into that discussion an explanation of the following three reasons why the AP (1986) test needs to be updated:

- First of all, on the order of four times as many daily stock returns observations currently available for U.S. firms – and we will here analyze data on companies of much greater current interest than those considered in the AP (1986) study, such as Pfizer, Ford, NVIDIA, and the like.
- Secondly, in view of the conditional heteroscedasticity well-known to be present in daily stock returns data, it is important to replace the ordinary bootstrap simulation methodology used in the AP (1986) computations, because the ordinary bootstrap assumes that the underlying data has a constant variance. Fortunately, all of the pre-whitened daily returns are symmetrically distributed, so this conditional heteroscedasticity issue can be readily handled using the ‘wild’ bootstrap.<sup>5</sup>

---

<sup>4</sup>Strictly speaking, this is a geometric quasi-random walk model – see Equation 1 below.

<sup>5</sup>As described in Section 2 below, we chose the Gonçalves and Kilian (2004) formulation of the wild bootstrap for use here, because it is optimized for the presence of both conditional heteroscedasticity and serial correlation in the daily returns data.

– Lastly, we now find that the usage of the word ‘test’ – and associated phrases, such as “we can reject the null hypothesis of serial independence at the 5 percent level” – in AP (1986) warrants a bit of caution. Or perhaps it would be more accurate to say that we are expanding on the way in which the AP (1986) test results can be most constructively described and interpreted. We will return to this topic at the close of Section 2.

There is nothing incorrect in the AP (1986) statistical setup, or in the usage there of the ‘testing’ nomenclature alluded to above. The AP (1986) formulation is indeed testing a specific null hypothesis – to the effect that a stock’s returns are, as a time-series, serially independent. And (as will be described in Section 2), the tail area of the (simulated) distribution of the maximal absolute sample excursions of the deviations of a stock’s  $\log(P_t)$  from its trend line does indeed – under the random walk null hypothesis of serially independent stock returns – correspond to the probability of observing evidence against the null hypothesis this strong (or stronger) due to chance. So it actually is correct to call this probability the ‘rejection P-value’ with respect to this null hypothesis of serial independence in the stock returns. Thus, a small P-value is indeed precisely this kind of formal measure of the strength of the empirical evidence against the proposition that this stock’s returns are serially independent.

On the other hand, we will also argue in Section 2 that this usage of the hypothesis testing framework – its verbiage and (in particular) its focus on a rejection P-value concept, wherein a P-value less than 0.05 brooks large – is, perhaps not the most informative way to use the information in this simulated distribution of the maximal absolute excursions of the stock log-prices from their trend line.

- In Section 3 we set forth and interpret the full set of new test results which we have obtained in the present work – beyond those for the 3M Company already ‘leaked’ to the reader as part of the illustrative example in Section 2. This full set of results is

based on sample daily returns data (from 1962 through 2023) on thirteen individual firms, as well as on both the CRSP value-weighted and individually-weighted stock indexes.<sup>6</sup>

Some of the results obtained and discussed in AP (1986) hold up quite well in the present study, even with the larger samples and better simulation methodology available here; others – e.g., our particular conclusions about the CRSP indexes – are updated in Section 3.

Section 3 concludes with a discussion of the numerous results we obtained, detailing how the results vary (for each stock) across nine, approximately-equal, 7-year-long subperiods. These subperiod-specific results, clearly, were only feasible because our new sample period was ca. three times longer than that available for the AP (1986) study.

- We conclude, in Section 4, with a summary of the overall significance of our new results using this updated version of the AP (1986) test – i.e., a bit about what we think that our current results mean – and go on to include a discussion as to some directions which our results suggest for future research.

## **2. The Ashley and Patterson (1986) Test for Serial Independence in Daily Stock Returns**

When Ashley and Patterson (1986) was written, it was commonplace to take a geometric random walk (around the trend) as the paradigmatic statistical representation of the daily returns to a share of stock in a U.S. corporation. As will be made more explicit below, this model for stock returns posits that the daily return in day  $t$  – i.e., the change each day in

---

<sup>6</sup>Table 1 lists the firms, their precise sample periods, and a few descriptive summary statistics for each. There the reader will note that this full sample length of ca. 15,000 days is not available for several of the more newly-listed companies, such as Tesla and NVIDIA. CRSP is the Center for Research in Security Prices, at url <https://www.crsp.org/research/crsp-us-stock-databases/>.

*The Ashley And Patterson (1986) Test For Serial Independence in Daily Stock Returns, Revisited*  $\ln(P_t)$ , the logarithm of the stock's price – is a constant plus a sequence of independently distributed random variates.

AP's contribution (in a nutshell) was to become concerned – in a constructive way – regarding the observation that this random walk assumption implies that the deviation of  $\ln(P_t)$  from its linear trend line will therefore, in magnitude, wander off to infinity over an extended period of time. In other words, as the sample length becomes unboundedly large, this random walk assumption implies that the magnitude of the deviation of  $\ln(P_t)$  from its trend line will become unboundedly large, even though the average value of this deviation from the trend line is, of course, zero.

This implication of the random walk assumption struck us – indeed, continues to strike us – as counterintuitive.<sup>7</sup> The 'unboundedly negative' part of this implication with regard to  $\ln(P_t)$  can, with fair grace, be interpreted as the firm eventually going bankrupt, and either disappearing altogether, or being swallowed up by another firm. The 'unboundedly positive' part, however, remains for us conceptually problematic.

What made this observation of ours of scientific value is that we proposed an observationally falsifiable hypothesis based upon it: the random walk hypothesis implies that an empirically-observable sequence of  $T$  daily stock returns should, we conjectured, 'wander off' from the trend line at least as far – and as often – as one would 'statistically expect' based on this hypothesis. But, crucially, it should not wander off more than this.

Thus, if this null hypothesis (positing that the geometric random walk model adequately describes the generating mechanism for daily stock returns) is *false* – say, because some nonlinear mechanism kicks in to 'pull' returns back toward the trend line whenever they deviate too far away from it – then one ought to find that the observed sequence of daily stock returns 'wanders off' from the trend line less far (or less frequently) than one might

---

<sup>7</sup>We note that pronouns such as 'we,' 'us,' and 'our' – here and below – obviously now include Najafi, who was born in 1982.

‘statistically expect’ it to, based on the random walk null hypothesis. Thus, the possibility of a usable scientific (one-sided) hypothesis test arises.

As noted in Section 1 above, the present work will improve on AP (1986) in several respects. In particular, we will now be able to utilize much longer samples on stock returns for U.S. firms than were available in 1986. And we will also here use data on a sample of firms which is both more numerous and more likely of current interest. In addition, we will also modernize the simulation technique used in constructing the test, to allow for the conditional heteroscedasticity in the daily stock return data; this was not possible at the time AP (1986) was written. And – further – we both hope and expect that the reader will find substantial additional value in our updated interpretation here of what we reported (and interpreted) in AP (1986) solely as hypothesis test rejection P-values. All of these improvements will be addressed (and further discussed) below.

In contrast, we see no way to materially improve upon the literature review and the description of the test in the Introduction and Methods sections of AP (1986, pp. 221-224). We therefore incorporate that material here – ‘by reference’, as it were – and direct the reader to the relevant JSTOR link, at <https://www.jstor.org/stable/2330739>.<sup>8</sup> On the other hand, a nice, graphics-laden illustrative example – better than was in AP (1986) – does certainly seem called for here.

So let’s now explicitly consider the application of the AP (1986) test to the daily returns (exclusive of dividends) on the stock in one of the companies we selected for analysis in the current work: the 3M Company.<sup>9</sup> These returns – denoted  $r_t$  here, to conform with the notation in AP (1986) – were downloaded from the CRSP database for the 15,603 day long sample period running from January 1, 1962 though December 31, 2023.<sup>10</sup>

---

<sup>8</sup>Access to this article on JSTOR does require a subscription, but it does not involve going to the library, or even arising from one’s office chair.

<sup>9</sup>This stock was selected for use in this example entirely because it was listed at the top of Table 1.

<sup>10</sup>See Tables 1 and 2. As noted earlier, CRSP is the Center for Research in Security Prices, at url <https://www.crsp.org/research/crsp-us-stock-databases/>.

Based on the geometric quasi-random walk model, then, the dynamics of the 3M Company share price itself –  $P_t$  – is implied by the equation:

$$(1) \quad \ln(P_t) = \mu + \ln(P_{t-1}) + r_t$$

where  $\mu$  is the slope of the log-linear trend line. The string ‘quasi’ is pre-pended above to allow for the possibility that the error term –  $r_t$  – in this model, which are the daily returns to the stock, might be serially correlated. This is the case for about half of the daily stock returns series which we have looked at, and is in fact the case for the 3M Company returns. But the amount of serial correlation we observe is always quite minor, generally yielding estimated AR(p) pre-whitening models with  $R^2$  values less than 0.05.

A time plot of the 3M Company daily returns observations over this 1962 - 2023 sample period is displayed as Figure 1 below; Figure 2 displays a histogram of these data, with a corresponding Gaussian density plot superimposed for comparison. These plots show that – as is typical of daily returns data – the 3M Company returns at first glance appear to be normally distributed and covariance stationary. On closer inspection of the histogram, however, one can also clearly see that these returns are leptokurtic – more sharply peaked and thicker-tailed than would be data drawn from a Gaussian distribution. And, on closer inspection of the time plot, one can clearly see the time-dependent bursts of higher variance typical of a time series exhibiting strong evidence of conditional heteroscedasticity. Thus, the 3M Company returns at first glance appear to be Gaussian and covariance stationary, but are actually neither, as is the case with all of the other returns data we consider here.<sup>11</sup>

---

<sup>11</sup>The reader might also note the presence of what appears to be a possible outlier in the 3M Company returns at day number 6,485, which corresponds to October 19, 1987. It is easy to eliminate this observation from the sample, which yields a time plot (and histogram) which both look a bit ‘nicer.’ The estimated pre-whitening model is little affected by this elimination. More importantly, the plot of the cumulative sum of the pre-whitened returns – discussed below – is almost completely unaffected by this elimination, so that the AP test result quoted in Table 2 for this stock changes only from 0.059 to 0.060. The AP test is evidently very robust to outliers.

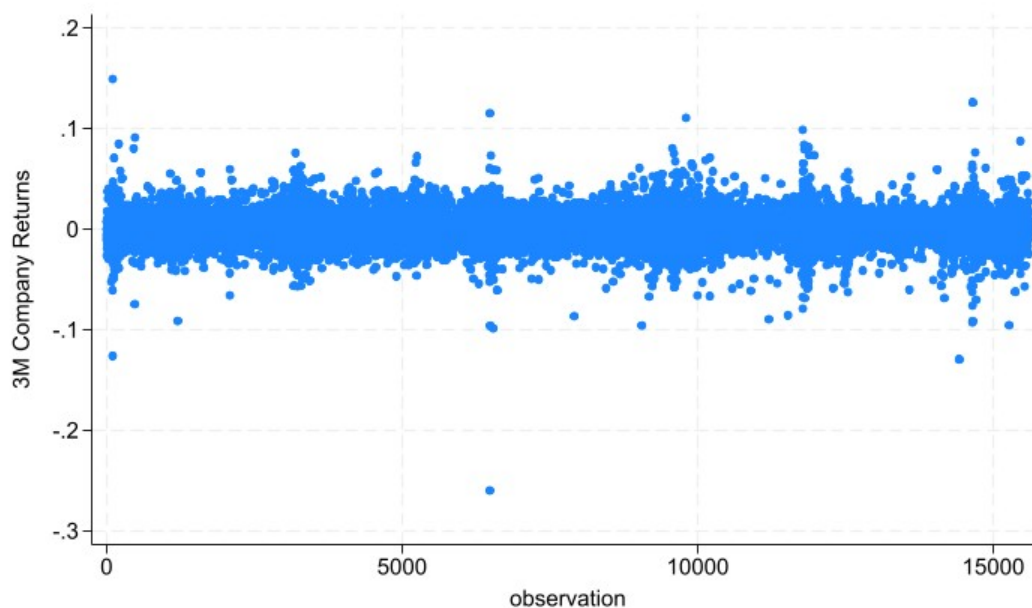


FIGURE 1. 3M Company daily returns: Time Plot January 1, 1962 through December 31, 2023

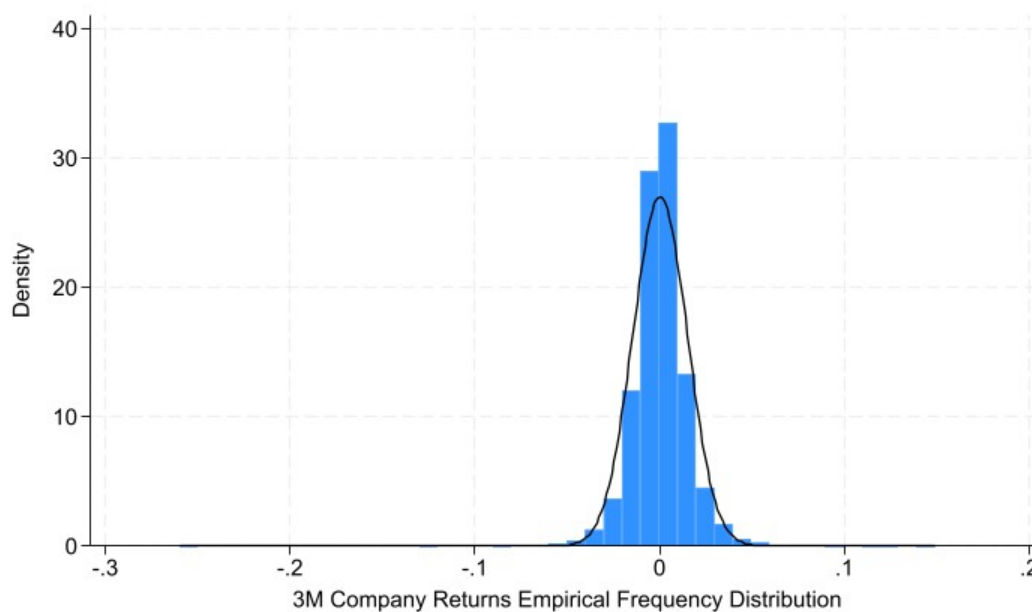


FIGURE 2. 3M Company daily returns: Histogram January 1, 1962 through December 31, 2023

These 3M Company daily returns are very slightly autocorrelated, so we specified and (using OLS) fit an AR(2) model to them, yielding the essentially uncorrelated fitting errors,  $x_3 \dots x_{15606}$ .<sup>12</sup> It would be pointless to also display a time plot of the  $x_t$  time series, because this AR(2) model has an  $R^2$  of only 0.0007, so this time plot would be visually indistinguishable from the time plot of the returns themselves.

We are also not displaying plots of either the sample autocorrelations or the partial autocorrelations here – for either the returns themselves or for the pre-whitening model fitting errors – because we did not find these plots very useful here. This is because the non-Gaussianity and conditional heteroscedasticity in these data (for 3M Company and for the rest of the stocks) violate the assumptions underlying the derivations of the asymptotic 95% confidence intervals for both of these diagnostic tools, rendering these sample plots essentially useless for specifying the value of  $p$ . Instead – in view of the large size of  $T$  and the fact that OLS still provides consistent estimates of the AR( $p$ ) model coefficients – our modeling strategy was to simply, when in doubt, somewhat overfit the AR( $p$ ) models: Because of the aforementioned consistency argument, there is no harm done if the value of  $p$  is set slightly too large; whereas we want to give the AR( $p$ ) model every chance to ‘mop up’ whatever linear mean-reversion mechanism might exist before ascribing any empirical features in the data to deficiencies in the serial independence assumption of the geometric random walk model.<sup>13</sup>

---

<sup>12</sup>Two fitting errors are lost due to the lags in the AR(2) model.

<sup>13</sup>Thus, for example, the coefficient at lag four in the Exxon pre-whitening model has an estimated t ratio of just 1.87, but we cheerfully retain it in the model specification, yielding an AR(4) model for the Exxon stock. Since this coefficient is consistently estimated, we feel that this estimated coefficient at lag four, while only marginally significant, is a better estimate than zero. We take this to be the ‘conservative’ choice, because allowing the pre-whitening model to be a bit stronger if anything reduces the empirical evidence against the random walk null hypothesis.

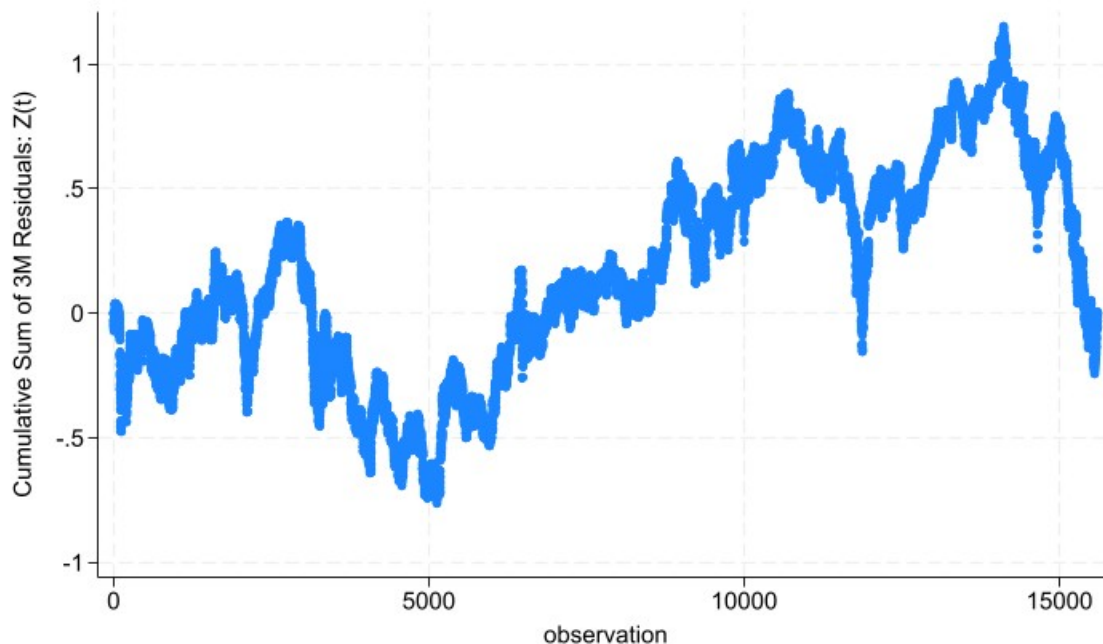


FIGURE 3. 3M Company Cumulated (Pre-Whitened) Daily Returns: January 1, 1962 through December 31, 2023

$Z_t$  is defined in AP (1986, p. 223) as the cumulative sum of these (uncorrelated, by construction) fitting errors. Which is to say,  $Z_t$  is the cumulative sum of the – now ‘pre-whitened’ – random walk steps:

$$(2) \quad Z_t = \sum_{j=1}^t x_j$$

For the (pre-whitened) 3M Company return data, this cumulative sum – the position of the random walk, as it were – is plotted in Figure 3.

Note that the time plot of  $Z_t$  for the 3M Company in Figure 3 must begin and end at zero, because OLS fitting errors must sum to precisely zero. And, of course, the  $Z_t$  time-series wanders smoothly around over the course of the sample – just as one would expect from its definition (in Equation 2) and from the random walk hypothesis embodied in Equation 1, both of which imply that  $Z_t$  is integrated of order one, i.e.,  $I(1)$ .

In the notation introduced in AP (1986), the largest observed deviation (in magnitude) of  $Z_t$  away from zero is denoted  $Z^{MAX}$  and is designated as the value of the statistic for the AP test.

In this case – of the 3M Company returns data – the maximum upward excursion of the  $Z_t$  random walk is 1.151 and the maximum downward excursion is -0.760, so  $Z^{MAX}$  equals 1.151. Thus, in this particular instance, the upward excursion is larger than was (in magnitude) the downward excursion. But we have generally observed no tendency (in the U.S. returns data which we have so far examined) for either the upward or the downward excursion to predominate in magnitude. In particular, in the five stocks (and two indexes) we considered in AP (1986), the maximum absolute excursion was positive in three cases and (including its sign) negative in four cases. And in the thirteen stocks (and two indexes) considered in the present study, the maximum absolute excursion was positive in eight cases and negative in seven cases.

The essential issue, then, is this: Is this observed  $Z^{MAX}$  value – the largest-magnitude excursion of  $Z_t$  away from zero – which is to say, the largest-magnitude cumulative sum of the (pre-whitened) returns – smaller than one would expect, under the null hypothesis that the random walk model is valid, so that the observed sample of (pre-whitened) returns  $x_t$  is in fact consistent with an independently distributed sequence of random variables?

If the observed  $Z^{MAX}$  value is ‘too small’, then AP (1986) argued that one could take this as empirical justification for rejecting the random walk hypothesis null hypothesis – of serially independent returns – in favor of a more nuanced story, in which some nonlinear ‘restoring force’ nudges daily returns back toward the trend whenever  $\ln(P_t)$  departs too far from its trend line.

In order to address this ‘too small’ question, AP (1986) simulated the sampling distribution of  $Z^{MAX}$ , using the ordinary (nonparametric) bootstrap to generate 5,000 simulated  $x_j$  values, leading to 5,000 simulated  $Z_t$  time-series – per Equation 2 – and 5,000 realizations of  $Z^{MAX}$ . The fraction of these realizations of  $Z^{MAX}$  which were less than what we will

henceforth denote as  $Z_{smp}^{MAX}$  was then entered by AP (1986) in their Table 1 under the column heading ‘Probability Value.’<sup>14</sup>

At this point two relevant observations ought to be addressed:

- First, the bootstrap methodology we used in AP (1986) took account of the non-Gaussian distribution of the pre-whitened returns –  $x_j$  in Equation 2 – but did not account for the evident conditional heteroscedasticity in the returns. Since that time, however, bootstrap simulation from a conditionally heteroscedastic (but serially independent) time series has become quite feasible – albeit only if the time series is symmetrically distributed; this can now be done using what is called the ‘wild’ bootstrap. Fortunately, in our application here – as is evident from the histogram for the 3M Company returns data displayed in Figure 2, and which is similar in this regard to the histograms for all of the other stock return time-series considered here – this symmetry requirement is not problematic.

A bit of background on bootstrap simulation: The ordinary bootstrap assumes that the original data are identically (as well as independently) distributed, and simply selects new – i.e., simulated – values by picking at random (with replacement) from the original sample. The wild bootstrap is different: it is aimed at reproducing in each generated time series whatever pattern of heteroscedasticity – conditional or not – existed in the original series. It does this by retaining the serial independence assumption of the ordinary bootstrap, but replacing the assumption that the data are identically distributed in each period with an assumption that the data are symmetrically distributed. The wild bootstrap then creates each ‘new’ sample observation by merely – at random – flipping the sign on this individual observation. Each new sequence of observations – i.e., each new sample – will therefore be completely different, but everything about its variance – including any pattern with which its

---

<sup>14</sup>As will explained shortly, this nomenclature is modified here, so that this quantity is denoted ‘*Frequency<sub>i</sub>*’ for firm  $i$  in the tables and text below.

variance evolves over time – is preserved in each of the newly generated samples. Clearly, this ‘sign flipping’ does not alter the distribution of the variate if and only if it is symmetric.

Thus, in principle, one can obtain a new sample from the wild bootstrap by just multiplying the original error term in the model (independently in each period) by any zero-mean, unit-variance variate that strikes one’s fancy. Most people, however, simply use for this purpose a simple Bernoulli random variable, that takes on the values of either plus or minus one.

Asymptotically, this distributional choice in the application of the wild bootstrap doesn’t matter. But in finite samples, it turns out that it does – especially in somewhat-fraught cases such as ours, where one has a toxic brew of both fairly strong conditional heteroscedasticity and (for a number of stocks) at least some degree of serial correlation. Fortunately for us, that is precisely the case which was extensively investigated by Gonçalves and Kilian (2004), whose extensive simulation work indicates that using realizations of an independently distributed unit normal to flip the signs works well in this setting; so that is the version of the wild bootstrap used here.

- Second, it is noteworthy that, when the pre-whitened 3M Company returns time series plotted in Figure 1 is simulated 10,000 times – using the wild bootstrap exactly as described above – we find that the 10,000 resulting simulated values of  $Z^{MAX}$  are usually larger than the observed value of  $Z^{MAX}$ ,  $Z_{smp}^{MAX}$ . In fact, we find that 94.1 percent of our bootstrap-generated values of  $Z^{MAX}$  – based on the independent-returns geometric random walk model embodied in Equation 1 – exceed the value of  $Z^{MAX}$  observed with the actual 3M Company returns,  $Z_{smp}^{MAX}$ , which equals 1.151.

Now, one way of expressing/interpreting this result – and the one that we were naturally inclined toward in AP (1986), when we obtained results not unlike this for stocks like Carolina Power & Light, Holly Sugar, etc. – is to conclude that the null

hypothesis of serial independence for the 3M Company stock returns can be rejected at the 6% level.

This conclusion is warranted here on two levels. In the first place, the rejection P-value here is  $1.000 - 0.941 = 0.059$ . More deeply – from a frequentist perspective – one could justify our calling this quantity ‘the rejection P-value’ by noting that, in fact, this is – on repeated sampling, based on this null hypothesis being valid – the actual frequency with which we observe (simulated) values of the test statistic ( $Z^{MAX}$ ) smaller than the value (1.151) observed in the sample data. Indeed, it is for this reason that we now, in Table 2 for the present work, enter this result (for the 3M Company returns) in a column with the heading ‘Frequency’.

And we continue to argue that this interpretation – in terms of a hypothesis test rejection P-value – is fundamentally correct. It is, in isolation, perhaps a trifle misleading, however. For one thing, this particular result, for the the 3M Company returns, is a bit atypical: for example, most of the rejection probabilities listed in Table 1 of AP (1986) were considerably larger than 0.05.

This issue was addressed in AP (1986) – and recurs in our results here – so it is discussed below in Section 3. There we describe the methods we used to explicitly quantify how atypical this 3M Company result is in our present results. Foreshadowing, we use here the same Fisher (1925) P-value combination method as in our AP (1986) paper, which requires an assumption that the test results for all of the stocks are statistically independent. This independence assumption is perhaps not seriously violated, but it is impossible to credibly check it – and it is more than a bit awkward that there is no way to ascertain how sensitive the validity joint test is to whatever violations are present. For this reason, we also in Section 3 apply the HMP P-value combination method – see Wilson (2019) – which is robust to statistical dependence between the test P-values. This HMP method thrives on combining larger numbers

of P-values, so we hoped that our sample of 13 firms here – in contrast to just 5 in AP (1986) might suffice.

In fact, the main impetus for the present update to AP (1986) was a hope on our part that the substantially larger samples (in terms of both the number of days and the number of firms) and the better bootstrap simulation methods available for the present study, would yield empirical rejections of the serial-independence random walk null hypothesis substantially more clear than the results which we were able to obtain in AP (1986). The degree to which we succeeded in that regard will become apparent in Section 3 below.

Before going on to those results, however, we think that it is essential at this point to describe how the present research effort has caused us to broaden and deepen the way we evaluate and interpret these test results.

A sensible way to interpret a rejection P-value is to posit that one either does – or does not – reject the null hypothesis, based upon the sample evidence against it. Then, upon repeated sampling – in a frequentist world, where one imagines that such repeated sampling is a possibility – and (notably) where the null hypothesis is actually true, the rejection P-value is just the fraction of the repeated samples where one (incorrectly) rejects the null hypothesis. This, as we have hopefully made clear above, is precisely what we are actually calculating with our bootstrap simulations, and quoting as the empirical frequency in the ‘*Frequency*’ columns of Tables 2 through 10 below.

At this point, however, we think it is useful and important to highlight a second, distinct interpretation of this empirical frequency, which we are here denoting as ‘*Frequency*’ or ‘*Frequency<sub>i</sub>*’ and also interpreting as the rejection P-value: one could alternatively interpret this empirical frequency simply as a monotonic, objective measure of the strength of the evidence against the null hypothesis, that the returns are serially independent.

On this interpretation, a ‘*Frequency*’ close to zero simply means that there is overwhelmingly strong sample evidence against the null hypothesis: the distribution of simulated maximum absolute excursions away from the trend line under the serial independence (random walk) assumption exhibits simulated returns wandering off away from the trend line to a greater extent across the 10,000 simulations than we observe in the actual 3M Company data. Whereas, a ‘*Frequency*’ approaching one half means that there is no empirical evidence whatever against the null hypothesis: the distribution of simulated maximum absolute excursions away from the trend line under the serial independence (random walk) assumption has simulated returns wandering off away from the trend line about as much as one might expect (in its median value) as is observed in the actual data 3M Company data.

From this point of view, there is clearly nothing special about a P-value of 0.05. Indeed – from this perspective – it was fundamentally a bit misleading to discuss the simulation-based calculation discussed above in terms of a ‘hypothesis test’ – with a ‘rejection P-value’ – in the first place: More meaningfully, what we have actually done here, with the pre-whitened 3M Company daily returns data, is that we have simulated 10,000 new returns time-series – based on the assumption that these returns are serially independent – and found that the serial independence assumption yields too much dispersion away from the trend in ca. 94% of our replications.

That, in our view, is quite strong evidence against the proposition of that the 3M Company returns are serially independent. Moreover, we think that most people would still regard it as strong evidence against this serial independence proposition, even if this percentage had turned out to be only 85% instead of 94%– and even though no one that we know thinks that a hypothesis test rejection at the 15% or 30% level means much of anything.

From this perspective, then, we invite the reader to re-frame the results of the AP ‘test’ in these terms, and to view the figures in the ‘*Frequency*’ column (in

each of the tables below) as “the estimated probability that data simulated data under the hypothesis that returns are serially independent, and hence will deviate from the trend line even more than the amount observed” – rather than only as “the simulation-based rejection P-value for null hypothesis that returns are serially independent” – even though these two statements are actually equivalent.<sup>15</sup>

- Furthermore: Intuitively – we interpret a small sample value of ‘*Frequency<sub>i</sub>*’ – i.e., a value close to zero – as empirical evidence that returns for stock in firm *i* show signs of having been materially affected by some sort of ‘restoring force’ – after having wandered (perhaps repeatedly) too far away from the trend during the course of the sample period. Whereas, we would interpret a large sample value of ‘*Frequency<sub>i</sub>*’ – i.e., a value close to one-half – as an absence of empirical evidence for any sign of materially significant impacts of such a ‘restoring force’ on the returns for stock in firm *i*. Perhaps this absence of evidence is because the restoring force is weak (or even nonexistent) for firm *i*. Or perhaps this absence is because the returns for stock *i* never (or too rarely) wandered sufficiently far away from the trend to be affected this restoring force.

In contrast, while it certainly is possible to observe a value for ‘*Frequency<sub>i</sub>*’ in the interval  $[0.50, 1.0]$ , it is not so clear what economic interpretation to place on such an observation, although physicists seem to have no compunctions about inventing bizarre repellent forces in like circumstances. A result of this nature would seem to us to arise instead from either sampling errors, the use of a sample which is seriously too small for this task, and/or fairly severe model mis-specification issues.<sup>16</sup> Fortunately,

---

<sup>15</sup>Note that both figures here – the 0.059 ‘*Frequency<sub>i</sub>*’ and the (equal) 0.059 ‘rejection P-value’ – objectively follow from the sample data; but they could easily be interpreted (subjectively) in a different way by two different analysts. This is precisely analogous to how two people might obtain the same OLS-estimated regression model, with the same  $R^2$  value; but one person might view this  $R^2$  value as adequate, whereas the other might deem it a poor fit.

<sup>16</sup>Swollen sampling errors are, of course, the natural consequence of torturing a sample of inadequate size. The category of model mis-specification subsumes fundamental errors in specifying the basic form of either the pre-whitening model and/or Equation 1 itself.

reference to Table 2 indicates that ‘*Frequency<sub>i</sub>*’ observations in this range are not the norm, at least until one starts poking into the subperiods, where they are still fairly unusual.

This concludes our description of the AP test, as updated here, via an application to the new sample data on daily returns to the 3M Company. In Section 3 below we next describe the rest of the new sample data set assembled for the present study. There we both report on and interpret the results of applying the updated test to these new data.

### 3. The Stock Returns Data Used Here and Our Test Results

**3.1. The Sample Data.** Table 1 concisely describes our present, substantially more extensive, data set, obtained – as was the sample data set in AP (1986) – from the CRSP website.<sup>17</sup>

We pursued no formal strategy in selecting the 15 daily return series listed in Table 1, but the following considerations applied:

- We felt that it was essential to again include both the value-weighted and the equally-weighted CRSP indexes (VWRETX and EWRETX) in the analysis.
- It also seemed essential to include Exxon (XOM) in the analysis, because this was the only one – of the other stocks in the AP (1986) group of firms – which survived for very long after the end of the sample period for that study.
- Beyond that, we wanted our current sample of firms to both exhibit a reasonable variety of firm sizes – as quantified, for example, by their market capitalizations – and to also exemplify a variety of different industries.

The reader might conceivably take the phrase ‘a variety of industries’ above to indicate that perhaps we were vitally interested in obtaining a group of 13 individual stock return time-series whose sample cross-correlations were so small that we could argue – as we did, to some extent, in AP (1986) – that one could use the standard

---

<sup>17</sup>Center for Research in Security Prices, *op. cit.*

Fisher (1925) result to combine them in a portmanteau test. That was not the case, however: We merely sought variety in our sample of firms, so that our results would be of greater interest to ourselves and to our prospective readers.<sup>18</sup>

- The message we hope that the reader takes from Table 1 is that this sample of firms – while in no objective sense ‘representative’ – is nevertheless quite a variegated and, we think, an interesting one.

We note that the mean return for each stock is tiny because these are daily returns; and that – more importantly – these figures vary quite a bit across the firms, as do the quoted market capitalization estimates for the firms.<sup>19</sup> The quoted sample variance estimate in the case of each firm also varies considerably across the firms; but the reader should be advised that the sample variance is not an unbiased estimator of the population (unconditional) variance for time-series like these, which are conditionally heteroscedastic.<sup>20</sup>

**3.2. The Full-Sample Results.** Table 2 presents our results using the the sample daily returns data – exactly as described in Section 2 above for the particular case of the pre-whitened daily stock returns for the 3M Company, only now for all 13 individual stocks in our sample. In addition, in the last two rows of this table, we present the analogous results analyzing the daily returns of the two aggregate CRSP stock indexes: VWRETX and EWRETX.

---

<sup>18</sup>In fact, the use of the Fisher (1925) combination result was somewhat fraught in AP (1986) – and remains so here – because it rests on an assumption that all of the individual tests are independently distributed from one another. We will return to this topic in Section 3.2 below, where results related to this independence assumption – including the Fisher (1925) P-value result – are briefly justified and discussed.

<sup>19</sup>These market capitalization estimates are quoted without specific dates; they are included here for rough comparison purposes only.

<sup>20</sup>The serial correlation in several of these time-series also causes bias in the sample variance, but (per Table 2) this serial correlation is likely too weak here to be consequential in this regard. In any case, any real investigation of the variances of these time-series would likely have involved estimating GARCH models for them, and been less informative than Table 1: the variance estimates here in this table are descriptive statistics merely.

Table 2 is labelled, ‘Full-Sample Test Results’ because we also partitioned the full sample into nine subperiods (each ca. 7 years in length), and analyzed the data for each stock in each of these subperiods. Those subperiod-specific results are collected in Tables 3 through 10, and these will be discussed separately, in Section 3.3.

Returning to the Table 2 results:

- First, note that about half – 8 out of 15 – of the daily returns time-series were found to be serially correlated. We therefore ‘pre-whitened’ each of these, using an  $AR(p)$  model, specified and estimated using OLS regression, which is known to provide consistent parameter estimation in this setting.<sup>21</sup> Note that, even where serial correlation is observed (and modeled) in any of these daily return time-series, it is exceedingly weak: the  $R^2$  values quoted in Table 2 for these estimated  $AR(p)$  models are all (except for the EWRETX model) notably less than 0.01.
- Second, how can the figures in the ‘Frequency’ column be interpreted?
  - If the null hypothesis of serial independence in the pre-whitened daily returns is true,<sup>22</sup> then we should expect the sample value of  $Z^{MAX}$  to be pretty close to the median of the distribution of  $Z^{MAX}$  simulated under the serial-independence null hypothesis. **In other words, sample evidence ‘in favor of the null hypothesis’ for firm  $i$  corresponds to observing a Frequency value for this firm close to 0.50.**
  - **In contrast, a Frequency value for firm  $i$  close to zero is evidence against the null hypothesis of serial independence in the pre-whitened daily returns for this firm.** In other words, this is evidence against the

---

<sup>21</sup>Recall, that the daily returns for each stock are being modeled in this way so as to eliminate any observed tendency for the time-series to ‘wander back’ toward the trend line to a larger degree than is implied by the serial independence null hypothesis – i.e., the random walk model of Equation 1 – due merely to un-modeled serial correlation.

<sup>22</sup>‘True’ – in case the reader needs a refresh – is essentially equivalent to the assertion that Equation (1) (positing that  $\ln(P_t)$  is taking a geometric random walk around a linear trend, with no ‘restoring force’ tending to push it back toward the trend line when it wanders well away from it) is actually an accurate depiction of how  $P_t$  is generated over time.

proposition that (pre-whitened)  $\ln(P_t)$  for this firm is taking a geometric random walk around a linear trend, with no ‘restoring force’ tending to push it back toward the trend line, when it wanders well away from it. Implicitly, this is also evidence in favor of such a restoring force.

**Note that it is because of the apparent existence of this restoring force in the real world that the median of the distribution of the maximal excursions ( $Z^{MAX}$ ) of the pre-whitened returns for firm  $i$  in the simulated (random walk) world – in which such forces are absent – is relatively large. *This* is why the distribution of the maximal excursions ( $Z^{MAX}$ ) – simulated under the random walk assumption – has so little area (probability mass, or *Frequency<sub>i</sub>*) at excursions smaller than the value of  $Z^{MAX}$  observed in the actual sample data,**

$$Z_{smp}^{MAX}.$$

– And, of course, it is this last effect: that – in the presence of sample evidence against the null hypothesis – the median of the simulated distribution of  $Z^{MAX}$  shifts to a higher value, making *Frequency<sub>i</sub>*, smaller, which yields the alternative – also correct – interpretation of the observed *Frequency<sub>i</sub>* value as the rejection P-value for the null hypothesis of serial independence in the pre-whitened daily returns. Many people then attach a substantial (apparently cultural) significance to whether or not this P-value is (or is not) smaller than 0.05.

- We turn now – again – to the figures in the Frequency column of Table 2. At first blush, the evidence in this column against the serial independence (random walk) null

---

<sup>23</sup>It is perhaps only a detail, but this might be a good point to remind the reader of the following point: With our larger sample here (of 15 firms), we were able to observe that the maximum sample excursion which the pre-whitened returns took during the course of the sample was a positive excursion for half of the firms and a negative excursion for half of the firms. This observed symmetry is why we here concentrate upon (and define)  $Z_{smp}^{MAX}$  as the maximum of the *absolute value* of the observed cumulative excursion of the pre-whitened returns during the course of the sample.

hypothesis – now that we have amassed a lot more data and considerably modernized our simulation methodology – does not appear to be all that strong.

In particular:

- Many of the fifteen values of  $Frequency_i$  which we display in Table 2 cluster considerably closer to 0.50 than to zero.
- And, on the ‘P-value’ interpretation, the analysis viewed as a statistical test almost rejects at the 5% percent level for the 3M Company and rejects – but only at the 8% level – for Exxon and Pepsi Cola. And this is out of 15 tests.<sup>24</sup>
- On the other hand, it is also noteworthy that 13 out of the 15  $Frequency_i$  values – i.e., all except the ones for Pfizer, Inc. (at 0.493) and Tyson Foods (at 0.576) – are substantially smaller than 0.40. That group of results seems pretty unlikely, doesn’t it?<sup>25</sup>
- Now, in point of fact, most readers might actually agree with the opinion expressed immediately above: “That group of results seems pretty unlikely, doesn’t it?” But the only way to go beyond merely expressing that opinion – and leaving it at that – is to deploy some established statistical technique for sensibly combining these  $Frequency_i$  results or, equivalently, P-values – to obtain a combined (portmanteau, if you will) P-value.

Obviously, this problem of combining test rejection P-values is hardly a new one, having been addressed with a fulsome and long-established literature, beginning with Fisher (1925). Much – but not all – of this literature assumes that all of the individual test results are statistically independent. Thus, in order to proceed any further,

---

<sup>24</sup>Moreover, we should further also note at this point that – in contrast to the apparent result in AP (1986) – in these new results the evidence against serial independence no longer appears to be materially stronger in the two aggregate indexes (VWRET<sub>X</sub> and EWRET<sub>X</sub>) than in the individual stocks.

<sup>25</sup>And what about the  $Frequency_i$  estimate of 0.576 for Tyson Foods, which actually exceeds 0.500? This, of course, is only a sample estimate, so one could always interpret its excess over the median as being due to sampling error. Otherwise, this result would seem to betoken some sort of model mis-specification, either in Equation 1, or in the way we are conceptualizing this analysis in terms of some sort of ‘restoring force’ when pre-whitened returns drift too far away from the trend. We will return to this point in Section 4.

we, at least initially, conditioned our analysis on an assumption that  $Frequency_1 \dots Frequency_{13}$  are all statistically independent of one another.<sup>26</sup> As will become apparent below, however, we did not stop there: we also applied the HMP combination method, a more recent P-value combination approach which allows for dependence in the test values.

- Returning to the combination methods that do require that the  $Frequency_i$  values are independent, however, the problem with this assumption is that it is both untestable and non-credible to an unusually high degree here. In particular, it is not possible to use the values of  $Frequency_1 \dots Frequency_{13}$  obtained via bootstrap simulation for this purpose, because the simulation methodology itself imposes statistical independence in these values across all of the replications.

Consequently, the best that we can do – in evaluating the credibility of this independence assumption with regard to the  $Frequency_i$  values across the firms – is to estimate the sample cross-correlations of the pre-whitened daily stock returns for the firms.

Unfortunately, the sample daily returns data at hand for doing this – while numerous – are distinctly unsuited for this task, as they are all notably both leptokurtic and conditionally heteroscedastic. In particular, out of the 78 possible distinct sample cross-correlation pairs, a good many were small in magnitude. But several were quite large in magnitude, and do not credibly appear to be statistically insignificant – ranging as high as  $0.342 \pm 0.003$  for  $\text{corr}(\text{NVDA}, \text{DUKE})$  and  $0.698 \pm 0.003$  for  $\text{corr}(\text{NIKE}, \text{DUKE})$ .<sup>27</sup> On the other hand, having estimated 78 cross-correlations,

---

<sup>26</sup>While none of the 13 individual firms is large compared to the market, it is well known that individual stock returns are related to the contemporaneous market return. We therefore did not want to assume that  $Frequency_{14}$  and  $Frequency_{15}$  – for VWRETX and EWRETX, respectively – were independently distributed from the rest, so – as in AP (1986) – we dropped these two market indexes from this part of the analysis and continue this section with just the 13 individual-firm stocks.

<sup>27</sup>These standard errors were estimated using the wild bootstrap (pairwise), under what would seem to be a reasonable assumption (in this particular context) that the two time-series are jointly independent, since we are here implicitly testing the precisely that null hypothesis.

one one should expect to find a few that appear to be statistically significant. Admittedly, these two cross-correlations seem extreme, but one could still wonder how impactful the corresponding population cross-correlations between these returns are on the corresponding relationships between the *Frequency<sub>i</sub>* values for these three firms. Overall, we would have to rate the credibility of the independence assumption – with regard to *Frequency<sub>1</sub> ... Frequency<sub>13</sub>* – as ‘Surely incorrect – like all econometric assumptions – but here of more-than usually-dubious’ validity.

- Therefore – as noted above – we present below two “portmanteau” rejection P-values – combining just the results (*Frequency<sub>1</sub> ... Frequency<sub>13</sub>*) on the 13 individual stocks – in which we assume that these 13 P-value results are statistically independent. And then we further also report below a third combination P-value which is an application of a more recently developed strand of this ‘test combination’ literature, which allows for dependence amongst the test results.

This newer method is called the ‘Harmonic Mean P-value’ or ‘HMP’ technique; it does not assume that the 13 *Frequency<sub>i</sub>* results are statistically independent. We did not discuss or use it in AP (1986) – in part because it was not available in those days, and in part because its effective use will ordinarily rest on combining a good deal more than the 5 stock returns we used in our 1986 study. In any case, we think it useful to describe the HMP approach a bit at this point, since the reader will likely find it substantially less familiar than the other two (independence-assumption-based) combination methods which we will apply here.

HMP methods approximate the true combined P-value as a weighted average of the  $L$  tests considered – with the accuracy of the approximation increasing as  $L$  increases.<sup>28</sup> Generically, these methods sacrifice test power in favor of allowing for

---

<sup>28</sup>In fact, equal weights are most commonly used. Readers who are steeped in the forecasting literature might well find this practice reminiscent of the Bates and Granger (1969) forecast combination approach, which also typically – and rather successfully – avoids explicitly dealing with the correlations between individual forecasts by simply averaging the forecasts together with equal weights.

any existing dependence amongst the individual tests. The technical details of the HMP formalism lie well beyond the scope of the present work; for a start on those, the reader can consult Wilson (2019). But we will note here that the basic idea behind the HMP method leverages a consideration of various partitionings of the  $L$  tests being combined; this, fundamentally, is why theoretical extensions to the approximate HMP portmanteau P-value can in fact be made exact as  $L$  becomes unboundedly large.

HMP-approximated P-values are generally conservative; that is, a practice of rejecting the null hypothesis whenever one obtains an HMP P-value less than 0.05 will cause one to wrongly reject the null hypothesis with probability less than 0.05 in repeated sampling, with this discrepancy becoming more marked as  $L$  increases unless one does a fairly sophisticated correction. Notably, Wilson (2019) provides a table illustrating these discrepancies as a function of  $L$ , in which the smallest  $L$  value listed is 10. It is thus not completely unreasonable to apply the HMP method here, with  $L$  equal to 13.<sup>29</sup>

- That said, our three different combined P-value results are as follows:
  - (1) (Deflated odds ratio) The joint probability of  $L$  independent events is simply the extended product of their individual probabilities. Here, however, it is surely reasonable to deflate the probability of each of these individual events – i.e.,  $Frequency_i$  – by 0.50 before multiplying these probabilities together. This deflation makes sense because a  $Frequency_i$  value of 0.50 corresponds to observing a maximal absolute excursion ( $Z^{MAX}$ ) at the median of its distribution (simulated under the null hypothesis), which would correspond to no evidence at all

---

<sup>29</sup>In contrast, the HMP method would likely not have been appropriate for the AP (1986) data set, where  $L$  was only 5.

for any restoring force toward the trend in the returns data for this firm.<sup>30</sup>

$$(3) \quad P_{odds \ ratio} = \prod_{i=1}^{13} \frac{Frequency_i}{0.50} = 0.000006$$

Even with this deflation, however,  $P_{odds \ ratio}$  is an unappealingly tiny number, although it appears to be less small if one places it in percentage terms.

On the other hand, it is certainly reasonable to think that these 13  $Frequency_i$  results are not quite independently distributed. Thus, perhaps the best way frame this result is to allow that  $P_{odds \ ratio}$  is a combination measure that is appears to be exquisitely sensitive to deficiencies in the independence assumptions.

(2) (Fisher's Combined Probability Test) Fisher (1925) provided the fundamental result in this area, showing that, for  $L = 13$  independent tests, with P-values,  $P_1 \dots P_L$ ,

$$(4) \quad -2 \sum_{i=1}^{13} \ln(Frequency_i) \sim \chi^2(26) = 42.077$$

where the  $Frequency_i$  values from Table 2 have been substituted for  $P_i$  in Equation 4 above.

This result in Equation 4 allows the joint null hypothesis – that the (pre-whitened) daily returns from all 13 stocks are serially independent – to be rejected with a combined P-value of  $P_{Fisher} = 0.0241$ , the area to the right of 42.077 for the  $\chi^2(26)$  density. In other words, on the Fisher test, the Random Walk null hypothesis can be rejected at the 3% level. We note that this result is not terribly different from that obtained in AP (1986), where the Fisher test – with 5 stocks and 2,800 - 4,900 daily observations – yielded a rejection at the 2% level. You might think we would do ‘better’ with so much more data; on the

---

<sup>30</sup>In essence, each of the factors in this extended product is a kind of odds ratio; hence, the subscript.

other hand we are now taking account of the conditional heteroscedasticity in the data.

In the interest of clarity, it might be worthwhile to briefly digress at this point so as to explicitly distinguish the two distinct usages of the word ‘independence’ in the previous paragraphs, so as to head off any possible confusion on this point:

- (a) The stock returns – and hence the P-values and the (equivalent) *Frequency<sub>i</sub>* values – are assumed (with dubious empirical support) to be independent across the  $L = 13$  firms here, in order for the Fisher (1925) test to be valid.
  - (b) It is the random walk null hypothesis – to the effect that the daily returns of these 13 stocks are serially independent on a daily basis – which is being tested here, and rejected at the 3% level on the Fisher (1925) test.
- (3) (Harmonic Mean P-value Combination Method) The HMP method was described in some detail – about as much as is appropriate here – above. It can be implemented quite simply – or in a really quite complex manner.<sup>31</sup> The HMP P-value computation implemented here is based on the very simplest version of the HMP test, which should yield the most conservative (largest) P-value. In this version one simply uses the equally-weighted average of the  $L$  individual P-values to be combined:

$$(5) \quad P_{HMP} = \frac{1}{13} \sum_{i=1}^{13} Frequency_i = 0.245$$

This result is not very encouraging if one is out to reject the null hypothesis, and find evidence for the ‘restoring force’ alluded to in Section 2. One could surely do better – in terms of obtaining a ‘stronger’ inference – by using the asymptotically exact version of the HMP test, in the R package ‘harmonicmeanp’ cited by Wilson (2019, p.1197). But we strongly doubt that using this version of the HMP test would make a huge difference with  $L$  this small: It simply

---

<sup>31</sup>See Wilson (2019) for details.

doesn't seem likely to us that this, rather technically sophisticated, version of the test would adjust the  $P_{HMP}$  result notably below, say 0.10 – or that we would find such a result both informative and credible if it did yield such a drastic adjustment, since the source of this improvement would be completely opaque to us.<sup>32</sup>

On the other hand, this initial HMP result is, in point of fact, intensely valuable here – even despite the large magnitude of our somewhat oversimplified  $P_{HMP}$  approximation – because our work with the HMP method yields very clear guidance on how to proceed. We will return to this point immediately below, and also in Section 4.

- Summarizing and interpreting these three full-sample results:
  - The  $P_{odds\ ratio}$  result looks to be so sensitive to what failures apparently do exist in the independence assumption with regard to  $Frequency_1 \dots Frequency_{13}$  as to yield an outlandishly strong rejection of the random walk null hypothesis.
  - In contrast (as in the AP (1986) paper) the Fisher (1925) Combined Probability Test result – while also based on this same independence assumption with regard to  $Frequency_1 \dots Frequency_{13}$  – appears to be nicely behaved. With our new data here, it yields a rejection at the 3% level of the joint null hypothesis that these 13 stock return time-series are all serially independent.
  - One might reasonably take this pair of results to indicate that we have obtained reasonably strong evidence against the random walk null hypothesis – and in favor of some sort of restoring force when a stock wanders ‘too far’ from its trend. Except that both of these test results combining the P-values (or, equivalently,

---

<sup>32</sup>It is efficient to exploit our colleagues' comparative advantages and – crucially – to learn from them. But we have rarely encountered an individual who can make highly skilled use of a statistical technique which is to them a total ‘black box.’ This is an observation – not a value judgement. We observe that the HMP waters deepen quite quickly; on the other hand, Wilson (2019) appears to provide a good entry point.

*Frequency<sub>i</sub>* values) rest on an assumption that these 13 test results are statistically independent of one another, an assumption which we are fairly certain – from the cross-correlation results – several of which were quoted above – is materially false. So perhaps it is a bit of a stretch to claim that we have obtained a clear rejection of this null hypothesis at the 3% – or even the 5% level: the evidence here is a bit too murky for that – just like in AP (1986).

- Nevertheless – as in AP (1986) – in view of our  $P_{odds\ ratio}$  and  $P_{Fisher}$  results, we have evidently obtained some substantial evidence against the random walk null hypothesis here: to the point where our statement above that ‘we have reasonably strong evidence against the null hypothesis’ seems a fair assertion for us to make. Perhaps that is the best that can be really said at the moment.
- In contrast – both foreshadowing Section 4 and looking to the future – the disappointing HMP combination test result – which is robust to any dependence between the *Frequency<sub>i</sub>* results for the stock returns on  $l = 1 \dots L$  individual firms – apparently has insufficient power (at least in its very simplest form, as we deployed it here) to effectively address the random walk null hypothesis with the present group of just  $L = 13$  stocks. However, as noted above, we seriously doubt that sophisticated implementations of the HMP testing method will sufficiently improve its performance (from  $P_{HMP} = 0.245$ ) as to yield useful inferential results, at least with the present sample of just 13 stocks.
- But this HMP result points the way forward: The HMP test combination methodology is known to thrive on larger values of  $L$ , the number of tests in the aggregation, whereas the CRSP database provides access to literally thousands of stock return series.

Moreover, we have found here that the computational burden of the AP (1986) test, as we have revised it here, is surprisingly quite modest. In particular, our current implementing computer code calculates *Frequency<sub>i</sub>* for the full sample

of data on the  $i^{th}$  firm – including the 10,000 bootstrap simulations – in only 80 seconds, using quite limited computing resources: an interpreted (rather than compiled) Stata script on a laptop computer. This run-time could thus likely be cut by an order of magnitude.

Thus, perhaps our most important result here is to discover that the next step in addressing/testing the random walk null hypothesis is pretty clearly to expand the number of stocks considered from  $L = 13$  (plus the two stock indexes) to more like  $L = 500$  or  $L = 1,000$  stocks.

This ends the discussion of the ‘Full-Sample Test Results.’ As noted at the beginning of this Section, we also analyzed the sample data for each stock in each of 9 subperiods. These results are described in the next Section.

**3.3. The Subsample Results.** The subperiod-specific results are collected in Tables 3 through 11, and also displayed in Figure 4. These results are described and explained as follows:

- We specified the subperiods – as listed in the Tables – *a priori*, using the following desiderata:
  - (1) Each subsample was to contain all of the sample data from an integer number of years.
  - (2) The subsamples were to, insofar as possible without wasting any data, be equal in length.
  - (3) The number of subperiods periods,  $m$ , was to be set as large as possible – ideally ca. 10 – so that this effort could yield an adequate picture of the degree to which  $Frequency_i$  varies across the subperiods for this  $i^{th}$  stock.
  - (4) The sample length in every subperiod should be as large as possible – ideally ca. 2,000 days – so as to not be too much smaller than the sample lengths (2,196 to 4,892 days) used in AP (1986).

- Based on these criteria, we chose to partition the entire sample – consisting of 62 years of daily observations – into 9 almost-equal subperiods, each consisting of the data on either 7 years (for the first 8 subperiods) or 6 years (for the last, or 9<sup>th</sup> subperiod).
- Looking at the results in Tables 3 to 10, a few observations immediately emerge for comment:
  - First, the reader will note that some cells in these tables are blank. This is because some firms did not begin operating until later in the full-sample period.
  - Second, note that the bottom row of each table repeats the full-sample results, for immediate comparison.
  - Third, the values of lag-order ( $p$ ) and the pre-whitening AR( $p$ ) model goodness-of-fit measure ( $R^2$ ) vary considerably across both the firms and the subperiods. But all of these pre-whitening models – like those estimated using the full sample on each firm – are still very weak.
  - Fourth (and most importantly) our initial hope that the  $Frequency_i$  results for each individual firm would be clearly stable across its subperiods was not borne out by the subsample results.
- In particular, looking at the results in Tables 3 to 10, we do see a substantial number of small  $Frequency_i$  estimates– some of them *quite* small – including XOM (subperiod #8), MMM (subperiods #4,6,7,8), PEP (subperiods #3,4,5,6,7,8,9), DIS (subperiods #3,5,6), TSN (subperiods #3,5) and DUK (subperiods #5,8,9). So there does seem to be some evidence in the subperiod results against the random walk null hypothesis. And, in entering the observed subperiod-specific results into the tables, it seemed to us that we perceived a pattern of smaller  $Frequency_i$  values in the earlier subperiods.
- But Tables 3 to 10 display so many results as to be more than a little indigestible. Consequently, we have summarized and displayed all of these  $Frequency_i$  observations in Figure 4 below.

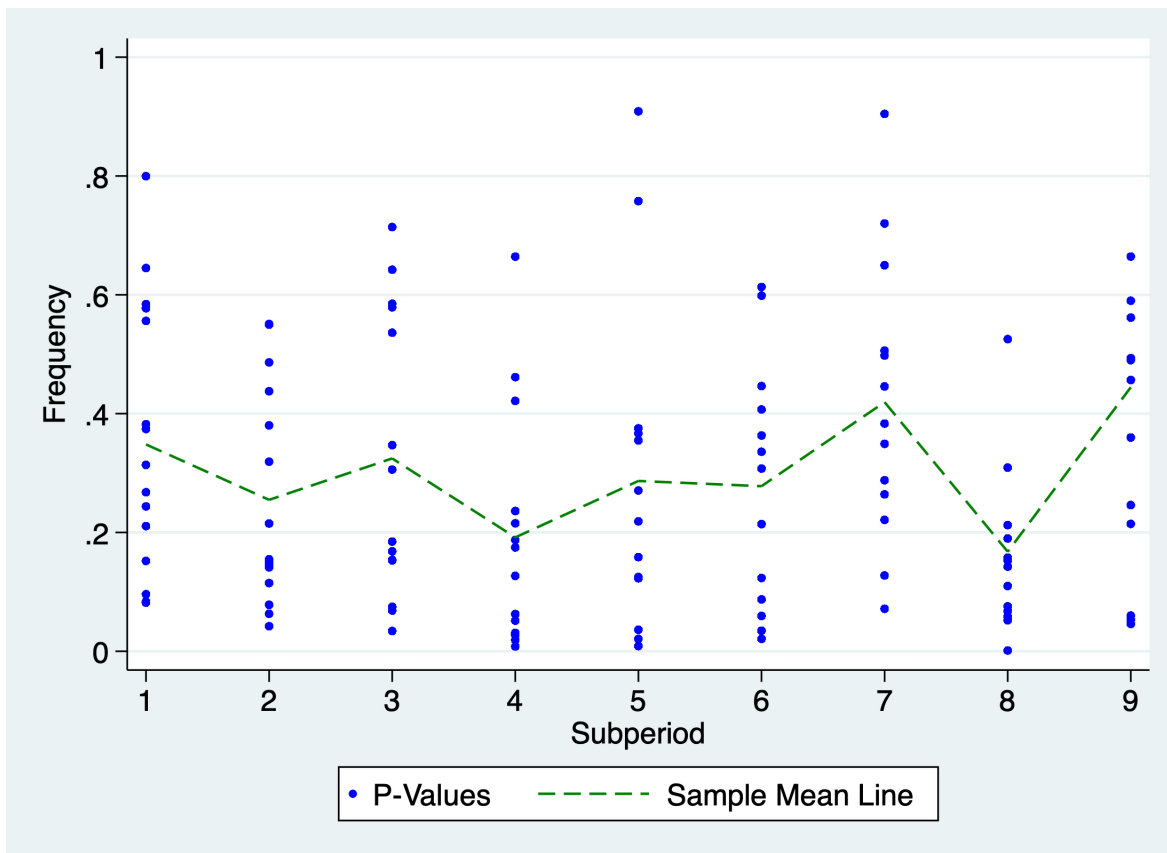


FIGURE 4. Subperiod-Specific  $Frequency_i$  Estimates Plotted vs. Subperiod

The structure of Figure 4 is very straightforward: Each dot in this figure plots one value of  $Frequency_i$ , for the subperiod which it is placed vertically above. And the dashed line in this Figure simply links the 9 subperiod-specific sample means of the  $Frequency_i$  observations.

Looking at Figure 4, any pattern that we thought we perceived in these subperiod-specific  $Frequency_i$  estimates was evidently in our minds only: appropriately plotted, it is difficult to see anything here except sampling error. Evidently, the lesson to be learned from our (considerable) efforts with these subperiod estimates – and it is an

important one – is that sample lengths this small are simply not useful in estimating  $Frequency_i$ .<sup>33</sup>

#### 4. A Brief Summary of the Results – and Our Views as to the Way Forward

Several of the authors have lived and worked in Blacksburg, Virginia for most of the time period spanned by AP (1986) to the present. This town lies at edge of a broad mountain valley, from which the mountain ridge line (while not outstandingly tall) is clearly visible. Until, that is, one initiates any sort of journey beyond the valley: at that point one soon loses sight of the mountains, and can only see the foothills.

In a similar vein, the preceding sections of this paper have examined the hills and dales of these foothills in all of the detail the reader might need.

In the present section, by way of contrast, we begin by singling out just the handful of ‘foothills-type’ results obtained here which we consider to be the most relevant and significant. Then we ‘crest the ridgeline’ – and close the paper – with our views as to the most constructive way forward in this research area.

##### (1) {Main New Results}

- Perhaps the most important thing we have learned in re-visiting the AP (1986) test of the random walk null hypothesis – Equation 1 – is how useful it is to ‘widen the aperture’ of how we were thinking about this topic. What we learned in this regard is not at all to abandon the AP (1986) procedure as a formal hypothesis test. Rather, we uncovered how to augment the meaningfulness of the P-value produced by the test by (in addition) re-framing and re-naming its realized value as the test statistic,  $Frequency_i$ .

---

<sup>33</sup>It is perhaps also worth adding that – as was noted earlier – it is difficult to interpret *population* values of  $Frequency_i$  larger than 0.50, but that comment hardly applies to sample estimates which apparently consist mainly of sampling errors.

$Frequency_i$  is still, of course, defined in terms of the simulated distribution of the maximum absolute excursions of the pre-whitened returns for stock  $i$ , which is explicitly specified in Section 2. But we think that it is valuable to note that  $Frequency_i$  – while numerically identical to the test rejection P-value – can additionally be interpreted as an empirical measure – to the degree to which it is close to zero – of the tendency for the  $i^{th}$  stock’s returns to not wander from the trend line as much as an assumption of serial independence would imply it would: as if there is some ‘restoring force’ pulling it back toward the trend line when it deviates too far from it.

- Using ca. 3 times as much sample data on ca. 2 times as many firms here – and now deploying ‘wild’ bootstrap simulation, so as to appropriately account for the conditional heteroscedasticity in the daily returns – we now have evidence against the (random walk) null hypothesis of serial independence in daily stock returns which is substantially more credible, statistically, than that which we were able to produce for the AP (1986) study. And, of course, our new results are also substantially more up-to-date.
- We must note, however, that our newly-gained statistical evidence against the null hypothesis of serial independence in these stock return series – while now more credible – is rather weak for the returns on each of the 13 firms individually.

On the other hand – viewing the 13 stocks as a group – the evidence on the 13 stocks is strongly suggestive of a ‘restoring force’: In fact, in Table 2 we find observed  $Frequency_i$  estimates which are notably less than 0.50 for 11 of the 13 stocks.

We therefore considered these 13 results as a group, using the Fisher (1925) Combined Probability Test to aggregate these 13 rejection P-values – i.e.,  $Frequency_i$  – results.

The Fisher test assumes that all 13 rejection P-values are statistically independent, however, an assumption which our results indicate is of distinctly dubious validity.<sup>34</sup> Still, combining these 13 individual-firm results from Table 2 in this way yields a rejection of the serial-independence null hypothesis at the 3% level, which we think is something of an achievement.<sup>35,36</sup>

## (2) Agenda For Future Work

We close with a couple of research topics suggested by the results obtained here:

- *{Expand the Sample to Eliminate the Independence Assumption}* The ability of the AP test to credibly reject the null hypothesis of serial independence in daily stock returns is hamstrung by the low power of the test. This low power is natural, in a way, because one should expect pre-whitened  $\ln(P_t)$  to only occasionally reach its maximum excursion away its trend, even over quite a long sample period. But the Fisher (1925) method for enhancing the power of the test, by combining P-values across tests for multiple stocks, relies on an unpalatable assumption that all of these test P-values are independent of one another. The HMP Combination Method avoids this across-firm independence assumption, but at the cost of low power and/or inaccuracy in its results – even the elaboration of it which purports to be asymptotically exact – unless and until one is able to include a quite sizeable number of P-values in the combination test.

Clearly, the constructive response here is to substantially increase the number of stocks considered, from the 5 stocks – in AP (1986) – and the 13 stocks considered

---

<sup>34</sup>See Section 3.2 for details.

<sup>35</sup>This test result is similar to the analogous 2% rejection using the Fisher method obtained in AP (1986), but we would not have the reader over-interpret this similarity, since AP (1986) uses a different sample of stocks and time periods, and a bootstrap simulation method not accounting for conditional heteroscedasticity in the returns series considered.

<sup>36</sup>We also obtained some not-so-useful results from a simplified application of the HMP P-value combination method, which – notably – does not require an independence assumption with regard to the 13 individual tests, *per* Wilson (2019). These results indicate that the HMP approach will require samples including *substantially* larger numbers of firms in order to be useful; see below.

here, to a sample of more like 500, or perhaps even 2,000 stocks. The requisite CRSP data is easily available. And our algorithm could easily be adapted to feasibly compute a large number of  $Frequency_i$  values, by simply re-coding it in a compilable computer language and running it on faster hardware.

- *{Use the AP Test as a PreTest to Enhance Stock Price Forecasting}*

Intuitively, rejection of the random walk null hypothesis considered here in the AP test – i.e., Equation (1), to the effect that pre-whitened daily stock returns are serially independent – in essence says that daily returns are forecastable from their own recent past, even beyond that provided by any weak short-term linear dependencies.<sup>37</sup> Further, it has also long been known that such rejections of serial independence additionally imply that daily stock returns can, in principle, be forecast using *nonlinear* time-series models. The problem has lain entirely in adequately specifying the form of such models.<sup>38</sup>

The opportunity here lies in the fact that a rejection of null hypothesis of serial independence on the AP test for a particular stock explicitly suggests something specific about why and how the random walk model is failing as a statistically adequate description of the process generating the returns for this stock: It is likely failing because some sort of nonlinear ‘restoring force’ comes into play whenever  $\ln(P_t)$  for this stock wanders too far away from its trend value.

If valid, this insight immediately suggests the specification of a 3-state *ESTAR* nonlinear model for any such stock – with a random walk specified for the middle state and a low-order  $AR(p)$  model specified for both outer states.<sup>39</sup> The degree to which the AP test is useful as a pre-test, in signalling when an *ESTAR* model

---

<sup>37</sup>Of course, the  $AR(p)$  pre-whitening models already indicate the existence of some linear forecastability in some stocks, but these models are so weak that this forecastability, in and of itself, is not of any practical utility.

<sup>38</sup>See Ashley(2012, pp. 632-637 and Enders (2014, Chapter 7) for more material on this topic of detecting and modeling nonlinear serial dependence in a time-series.

<sup>39</sup>See Ashley(*op. cit.*) and Enders (*op. cit.*) for details on *ESTAR* modeling.

can provide useful forecasting improvements is, we think, a question that is worth investigating.

- *{Endogenize Restoring Forces in Stock Price Theory}*
  - Finally, we think it reasonable to claim that our empirical work here goes quite some way toward establishing the empirical existence of the ‘restoring force’ – referred to repeatedly above – in at least some stock prices, whose relative strength for the stock of firm  $i$  is empirically (i.e., objectively) accessible by means of the  $Frequency_i$  statistic defined here. In other words, this characteristic now, we think, qualifies as an empirically-accessible ‘stylized fact’ about a firm.
  - The substantial inter-firm variation which we have observed in this  $Frequency_i$  statistic here suggests that it might well be highly informative to investigate the empirical sources of this variation. Ideally, this investigation would take place in the setting of a large panel data model. As usual, one hopes to source ideas for useful explanatory variables to include in this model from both informal and formal economic theory, a topic to which we will return shortly.

The research effort involved in constructing this panel data model might be closely related to that involved with the enhancement of the effectiveness of the HMP Combination Method (via increasing  $L$ ), alluded to above. Or it might not. For example, the dependent variable in such a model might more naturally involve  $\ln(P_{t,i})$  – rather than  $Frequency_{t,i}$ . Or it might be optimal to specify the time-periods for such a model on a much finer time-scale than the quite lengthy periods used here, since the underlying data are available on a daily basis.<sup>40</sup>

---

<sup>40</sup>Whoever takes on this project might also want give particular consideration to modeling  $\ln(P_{t,i})$  in levels rather than growth rates, and to using something akin to quantile regression estimation rather than OLS.

– Lastly – although we are not theorists – we dare to suggest to the theoretical community that it would be very useful – albeit perhaps quite challenging – to produce a dynamic model for the determination of the market price of a firm’s stock with strong microeconomic foundations (likely straddling both the world of point processes and the world of continuous/daily time), which is sufficiently rich as to be able to meaningfully confront and model data with variation in *Frequency<sub>i</sub>*. We envision such a theory as endogenously yielding something akin to the ‘restoring forces’ alluded to here, as a sort of emergent property. Any such theoretical model would be vitally useful, in both better understanding the sources of these restoring forces and in guiding empiricists in their specification choices for that nascent panel data model alluded to immediately above: Ideally, this scientific process is supposed to be a two-way street.

## REFERENCES

- [1] Ashley, Richard A., and D.M. Patterson. 1989. “Linear Versus Nonlinear Macroeconomies: A Statistical Test”. *International Economic Review* 30:685-704.
- [2] Ashley, Richard A. 2012. *Fundamentals of Applied Econometrics*. Wiley: New Jersey. ISBN 978-0-470-59182-6.
- [3] Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. *OR*, 20(4), 451–468. <https://doi.org/10.2307/3008764>
- [4] Enders, Walter 2014. *Applied Econometric Time Series*. Wiley: New Jersey. ISBN: 978-1-118-80856-6.
- [5] Fisher, R.A. 1925. *Statistical Methods for Research Workers*. Oliver and Boyd (Edinburgh). ISBN 0-05-002170-2.
- [6] Gonçalves, Silvia, and Lutz Kilian. 2004. “Bootstrapping Autoregressions With Conditional Heteroskedasticity Of Unknown Form”. *Journal of Econometrics*, 123, 89-120.
- [7] Wilson, D. J. 2019. “The harmonic mean p-value for combining dependent tests”. *Proceedings of the National Academy of Sciences USA*. 116 (4) 1195–1200. doi:10.1073/pnas.1814092116.

Table 1. Descriptive Statistics for the Stocks Used Here

| Company (Ticker)              | Market Cap | Sample Period            | T <sup>a</sup> | Mean     | Variance <sup>b</sup> |
|-------------------------------|------------|--------------------------|----------------|----------|-----------------------|
| 3M Company (MMM)              | 51 B\$     | 1962-01-02<br>2023-12-29 | 15603          | 0.000321 | 0.000218              |
| Duke Energy (DUK)             | 65 B\$     | 1962-01-02<br>2023-12-29 | 15602          | 0.000225 | 0.000192              |
| Exxon Mobil (XOM)             | 430 B\$    | 1962-01-02<br>2023-12-29 | 15601          | 0.000372 | 0.000213              |
| Ford Motor (F)                | 49 B\$     | 1962-02-02<br>2023-12-29 | 15610          | 0.000342 | 0.000454              |
| General Electric (GE)         | 183 B\$    | 1962-01-02<br>2023-12-29 | 15605          | 0.00036  | 0.000298              |
| Intl. Business Machines (IBM) | 180 B\$    | 1962-01-02<br>2023-12-29 | 15603          | 0.000326 | 0.000252              |
| Nike Inc (NIKE)               | 165 B\$    | 1980-12-02<br>2023-12-29 | 10855          | 0.000819 | 0.000462              |
| Nvidia Corp (NVDA)            | 2265 B\$   | 1999-01-22<br>2023-12-29 | 6275           | 0.001846 | 0.001441              |
| Pepsi Cola (PEP)              | 233 B\$    | 1962-01-02<br>2023-12-29 | 15605          | 0.000491 | 0.00023               |
| Pfizer Inc (PFE)              | 162 B\$    | 1962-01-02<br>2023-12-29 | 15618          | 0.000427 | 0.000285              |
| Tesla Inc (TSLA)              | 790 B\$    | 2010-06-29<br>2023-12-29 | 3398           | 0.002126 | 0.001285              |
| Tyson Foods (TSN)             | 25 B\$     | 1972-12-14<br>2023-12-29 | 12869          | 0.000775 | 0.000503              |
| Walt Disney Company (DIS)     | 165 B\$    | 1962-01-02<br>2023-12-29 | 15618          | 0.00067  | 0.000394              |
| Value Weighted Index (VWRETX) |            | 1962-01-02<br>2023-12-29 | 15605          | 0.000424 | 0.000101              |
| Equal Weighted Index (EWRETX) |            | 1962-01-02<br>2023-12-29 | 15605          | 0.000702 | 0.000078              |

<sup>a</sup>  $T$ , throughout the manuscript denotes the effective number of sample observations in the OLS AR(p) pre-whitening regression model, which excludes any missing observations. The slight variation in  $T$  for the samples beginning in 1962 thus also reflects differing holiday reporting of daily returns observations by the firms.

<sup>b</sup> This sample variance should be interpreted solely as a descriptive statistic, because daily stock returns exhibit conditional heteroskedasticity; therefore, the unconditional sample variance is not reliably estimated by the usual sample variance.

Table 2. Consolidated Full-Sample Test Results

| Company (Ticker)              | Sample Period            | T <sup>a</sup> | AR(p) <sup>b</sup><br>[R <sup>2</sup> ] | Frequency <sup>c</sup> |
|-------------------------------|--------------------------|----------------|-----------------------------------------|------------------------|
| 3M Company (MMM)              | 1962-01-02 to 2023-12-29 | 15603          | AR(2)<br>[0.001]                        | 0.059                  |
| Duke Energy (DUK)             | 1962-01-02 to 2023-12-29 | 15602          | AR(1)<br>[0.002]                        | 0.292                  |
| Exxon Mobil (XOM)             | 1962-01-02 to 2023-12-29 | 15601          | AR(4)<br>[0.007]                        | 0.080                  |
| Ford Motor (F)                | 1962-01-02 to 2023-12-29 | 15610          | AR(0)<br>[0.000]                        | 0.227                  |
| General Electric (GE)         | 1962-01-02 to 2023-12-29 | 15605          | AR(0)<br>[0.000]                        | 0.257                  |
| Intl Business Machines (IBM)  | 1962-01-02 to 2023-12-29 | 15603          | AR(0)<br>[0.000]                        | 0.145                  |
| Nike Inc (NIKE)               | 1980-12-02 to 2023-12-29 | 10855          | AR(3)<br>[0.003]                        | 0.184                  |
| Nvidia Corp (NVDA)            | 1999-01-22 to 2023-12-29 | 6275           | AR(0)<br>[0.000]                        | 0.288                  |
| Pepsi Cola (PEP)              | 1962-01-02 to 2023-12-29 | 15605          | AR(0)<br>[0.000]                        | 0.076                  |
| Pfizer Inc (PFE)              | 1962-01-02 to 2023-12-29 | 15618          | AR(2)<br>[0.003]                        | 0.493                  |
| Tesla Inc (TSLA)              | 2010-06-29 to 2023-12-29 | 3398           | AR(0)<br>[0.000]                        | 0.160                  |
| Tyson Foods (TSN)             | 1972-12-14 to 2023-12-29 | 12869          | AR(1)<br>[0.001]                        | 0.576                  |
| Walt Disney Company (DIS)     | 1962-01-02 to 2023-12-29 | 15618          | AR(2)<br>[0.002]                        | 0.342                  |
| Value Weighted Index (VWRETX) | 1962-01-02 to 2023-12-29 | 15605          | AR(0)<br>[0.000]                        | 0.260                  |
| Equal Weighted Index (EWRETX) | 1962-01-02 to 2023-12-29 | 15605          | AR(3)<br>[0.032]                        | 0.360                  |

<sup>a</sup> T, throughout the manuscript, denotes the number of sample observations.

<sup>b</sup> Maximum value of the cumulative sum of the fitting errors of the AR(p) model for the log returns of the company's stock return (excluding dividends) over the sample period. The absolute value of this number is the observed value of  $Z^{max}$ .

<sup>c</sup> The frequency (across the 10,000 bootstrap-simulated samples) of observing a simulated value of  $Z^{max}$  which is smaller than  $Z_{samp}^{max}$ , which is the value for  $Z^{max}$  observed in the actual sample data. Per the discussion in the text, this frequency can be interpreted as a rejection P-value or (if it is less than 0.50) simply as evidence against the null hypothesis of serial independence in the pre-whitened returns.

Table 3. Test Results for IBM and XOM by Subperiod

|                            | Intl. Business Machines (IBM) |       |           | Exxon (XOM)                |       |           |
|----------------------------|-------------------------------|-------|-----------|----------------------------|-------|-----------|
| Sample Period <sup>a</sup> | AR(p)<br>[R <sup>2</sup> ]    | T     | Frequency | AR(p)<br>[R <sup>2</sup> ] | T     | Frequency |
| Jan 2 1962 to Dec 31 1968  | AR(0)<br>[0]                  | 1737  | 0.577     | AR(2)<br>[0.008]           | 1734  | 0.369     |
| Jan 1 1969 to Dec 31 1975  | AR(1)<br>[0.006]              | 1761  | 0.141     | AR(1)<br>[0.036]           | 1765  | 0.155     |
| Jan 1 1976 to Dec 31 1982  | AR(1)<br>[0.007]              | 1766  | 0.536     | AR(0)<br>[0]               | 1769  | 0.306     |
| Jan 1 1983 to Dec 31 1989  | AR(1)<br>[0.004]              | 1768  | 0.422     | AR(4)<br>[0.039]           | 1765  | 0.190     |
| Jan 1 1990 to Dec 31 1996  | AR(0)<br>[0]                  | 1771  | 0.758     | AR(4)<br>[0.029]           | 1767  | 0.368     |
| Jan 1 1997 to Dec 31 2003  | AR(0)<br>[0]                  | 1761  | 0.336     | AR(5)<br>[0.016]           | 1756  | 0.110     |
| Jan 1 2004 to Dec 31 2010  | AR(0)<br>[0]                  | 1763  | 0.383     | AR(2)<br>[0.057]           | 1761  | 0.290     |
| Jan 1 2011 to Dec 31 2017  | AR(0)<br>[0]                  | 1761  | 0.158     | AR(3)<br>[0.006]           | 1761  | 0.067     |
| Jan 1 2018 to Dec 29 2023  | AR(0)<br>[0]                  | 1509  | 0.214     | AR(0)<br>[0]               | 1509  | 0.661     |
| Jan 2 1962 to Dec 29 2023  | AR(0)<br>[0]                  | 15603 | 0.146     | AR(4)<br>[0.007]           | 15601 | 0.085     |

The first sample observation in the CRSP data set is January 2, 1962. In a departure from a perfect level of consistency, we list the beginning and ending dates of the sample periods for each of the remaining eight subperiods in this table (and those to follow), so as to convey to the reader that we chose all of these subperiods – except, of course, the last – to be two calendar years in length. We have no idea why there are ca. 30 fewer days in the CRSP record between 1962 and 1968.

Table 4. Test Results for F and GE by Subperiod

|                              | Ford Motor (F)       |       |           | General Electric Company (GE) |       |           |
|------------------------------|----------------------|-------|-----------|-------------------------------|-------|-----------|
| Sample Period                | $AR(p)$<br>[ $R^2$ ] | T     | Frequency | $AR(p)$<br>[ $R^2$ ]          | T     | Frequency |
| Jan 2 1962 to<br>Dec 31 1968 | $AR(0)$<br>[0]       | 1735  | 0.211     | $AR(0)$<br>[0]                | 1736  | 0.096     |
| Jan 1 1969 to<br>Dec 31 1975 | $AR(1)$<br>[0.009]   | 1764  | 0.115     | $AR(0)$<br>[0]                | 1766  | 0.102     |
| Jan 1 1976 to<br>Dec 31 1982 | $AR(0)$<br>[0]       | 1769  | 0.579     | $AR(0)$<br>[0]                | 1769  | 0.347     |
| Jan 1 1983 to<br>Dec 31 1989 | $AR(4)$<br>[0.005]   | 1765  | 0.462     | $AR(0)$<br>[0]                | 1769  | 0.008     |
| Jan 1 1990 to<br>Dec 31 1996 | $AR(1)$<br>[0.000]   | 1770  | 0.355     | $AR(0)$<br>[0]                | 1771  | 0.273     |
| Jan 1 1997 to<br>Dec 31 2003 | $AR(0)$<br>[0]       | 1762  | 0.308     | $AR(0)$<br>[0]                | 1761  | 0.407     |
| Jan 1 2004 to<br>Dec 31 2010 | $AR(0)$<br>[0]       | 1763  | 0.904     | $AR(0)$<br>[0]                | 1763  | 0.650     |
| Jan 1 2011 to<br>Dec 31 2017 | $AR(0)$<br>[0]       | 1763  | 0.153     | $AR(0)$<br>[0]                | 1761  | 0.525     |
| Jan 1 2018 to<br>Dec 29 2023 | $AR(0)$<br>[0]       | 1512  | 0.563     | $AR(0)$<br>[0]                | 1509  | 0.493     |
| Jan 2 1962 to<br>Dec 29 2023 | $AR(0)$<br>[0]       | 15610 | 0.229     | $AR(0)$<br>[0]                | 15605 | 0.252     |

Table 5. Test Results for MMM and NIKE by Subperiod

|                              | 3M Company (MMM)           |       |           | Nike Inc (NIKE)            |                  |           |
|------------------------------|----------------------------|-------|-----------|----------------------------|------------------|-----------|
| Sample Period                | AR(p)<br>[R <sup>2</sup> ] | T     | Frequency | AR(p)<br>[R <sup>2</sup> ] | T                | Frequency |
| Jan 2 1962 to<br>Dec 31 1968 | AR(0)<br>[0]               | 1736  | 0.152     |                            |                  |           |
| Jan 1 1969 to<br>Dec 31 1975 | AR(1)<br>[0.032]           | 1765  | 0.078     |                            |                  |           |
| Jan 1 1976 to<br>Dec 31 1982 | AR(0)<br>[0]               | 1769  | 0.153     | AR(1)<br>[0.053]           | 527 <sup>a</sup> | 0.170     |
| Jan 1 1983 to<br>Dec 31 1989 | AR(0)<br>[0]               | 1771  | 0.027     | AR(1)<br>[0.009]           | 1771             | 0.572     |
| Jan 1 1990 to<br>Dec 31 1996 | AR(0)<br>[0]               | 1771  | 0.123     | AR(1)<br>[0.005]           | 1770             | 0.164     |
| Jan 1 1997 to<br>Dec 31 2003 | AR(2)<br>[0.124]           | 1759  | 0.060     | AR(0)<br>[0]               | 1761             | 0.183     |
| Jan 1 2004 to<br>Dec 31 2010 | AR(0)<br>[0]               | 1763  | 0.506     | AR(0)<br>[0]               | 1763             | 0.151     |
| Jan 1 2011 to<br>Dec 31 2017 | AR(0)<br>[0]               | 1761  | 0.076     | AR(3)<br>[0.009]           | 1758             | 0.156     |
| Jan 1 2018 to<br>Dec 29 2023 | AR(0)<br>[0]               | 1509  | 0.060     | AR(0)<br>[0]               | 1509             | 0.276     |
| Jan 2 1962 to<br>Dec 29 2023 | AR(2)<br>[0.007]           | 15603 | 0.056     | AR(3)<br>[0.003]           | 10855            | 0.19      |

T is smaller here because the sample return available for Nike was in for December 12, 1980. In general, see Table 1 or 2 for precise sample sample starting dates.

Table 6. Test Results for NVDA and PFE by Subperiod

|                              | Nvidia Corporation (NVDA) |      |           | Pfizer Inc (PFE)     |       |           |
|------------------------------|---------------------------|------|-----------|----------------------|-------|-----------|
| Sample Period                | $AR(p)$<br>[ $R^2$ ]      | T    | Frequency | $AR(p)$<br>[ $R^2$ ] | T     | Frequency |
| Jan 2 1962 to<br>Dec 31 1968 |                           |      |           | $AR(0)$<br>[0]       | 1743  | 0.084     |
| Jan 1 1969 to<br>Dec 31 1975 |                           |      |           | $AR(1)$<br>[0.014]   | 1772  | 0.149     |
| Jan 1 1976 to<br>Dec 31 1982 |                           |      |           | $AR(0)$<br>[0]       | 1769  | 0.034     |
| Jan 1 1983 to<br>Dec 31 1989 |                           |      |           | $AR(0)$<br>0         | 1769  | 0.019     |
| Jan 1 1990 to<br>Dec 31 1996 |                           |      |           | $AR(4)$<br>[0.007]   | 1771  | 0.219     |
| Jan 1 1997 to<br>Dec 31 2003 | $AR(0)$<br>[0]            | 1241 | 0.314     | $AR(2)$<br>[0.010]   | 1759  | 0.540     |
| Jan 1 2004 to<br>Dec 31 2010 | $AR(3)$<br>[0.007]        | 1760 | 0.144     | $AR(2)$<br>[0.011]   | 1761  | 0.356     |
| Jan 1 2011 to<br>Dec 31 2017 | $AR(0)$<br>[0]            | 1761 | 0.714     | $AR(0)$<br>[0]       | 1761  | 0.143     |
| Jan 1 2018 to<br>Dec 29 2023 | $AR(0)$<br>[0]            | 1510 | 0.218     | $AR(0)$<br>[0]       | 1509  | 0.490     |
| Jan 2 1962 to<br>Dec 29 2023 | $AR(0)$<br>[0]            | 6275 | 0.285     | $AR(1)$<br>[0.003]   | 15617 | 0.498     |

Table 7. Test Results for PEP and TSLA by Subperiod

|                              | Pepsi Cola (PEP)     |       |           | Tesla, Inc. (TSLA)   |      |           |
|------------------------------|----------------------|-------|-----------|----------------------|------|-----------|
| Sample Period                | $AR(p)$<br>[ $R^2$ ] | T     | Frequency | $AR(p)$<br>[ $R^2$ ] | T    | Frequency |
| Jan 2 1962 to<br>Dec 31 1968 | $AR(0)$<br>[0]       | 1736  | 0.244     |                      |      |           |
| Jan 1 1969 to<br>Dec 31 1975 | $AR(1)$<br>[0.023]   | 1765  | 0.438     |                      |      |           |
| Jan 1 1976 to<br>Dec 31 1982 | $AR(1)$<br>[0.011]   | 1768  | 0.069     |                      |      |           |
| Jan 1 1983 to<br>Dec 31 1989 | $AR(3)$<br>[0.012]   | 1766  | 0.052     |                      |      |           |
| Jan 1 1990 to<br>Dec 31 1996 | $AR(2)$<br>[0.006]   | 1769  | 0.036     |                      |      |           |
| Jan 1 1997 to<br>Dec 31 2003 | $AR(1)$<br>[0.008]   | 1761  | 0.021     |                      |      |           |
| Jan 1 2004 to<br>Dec 31 2010 | $AR(0)$<br>[0]       | 1763  | 0.072     | $AR(0)$<br>[0]       | 128  | 0.238     |
| Jan 1 2011 to<br>Dec 31 2017 | $AR(0)$<br>[0]       | 1761  | 0.001     | $AR(0)$<br>[0]       | 1761 | 0.396     |
| Jan 1 2018 to<br>Dec 29 2023 | $AR(1)$<br>[0.044]   | 1508  | 0.053     | $AR(0)$<br>[0]       | 1509 | 0.333     |
| Jan 2 1962 to<br>Dec 29 2023 | $AR(0)$<br>[0]       | 15605 | 0.076     | $AR(0)$<br>[0]       | 3398 | 0.178     |

Table 8. Test Results for DIS and TSN by Subperiod

|                              | Walt Disney Company (DIS) |       |           | Tyson Foods Inc (TSN) |       |           |
|------------------------------|---------------------------|-------|-----------|-----------------------|-------|-----------|
| Sample Period                | $AR(p)$<br>[ $R^2$ ]      | T     | Frequency | $AR(p)$<br>[ $R^2$ ]  | T     | Frequency |
| Jan 2 1962 to<br>Dec 31 1968 | $AR(1)$<br>[0.010]        | 1742  | 0.556     |                       |       |           |
| Jan 1 1969 to<br>Dec 31 1975 | $AR(1)$<br>[0.006]        | 1771  | 0.380     | $AR(2)$<br>[0.020]    | 764   | 0.414     |
| Jan 1 1976 to<br>Dec 31 1982 | $AR(1)$<br>[0.017]        | 1770  | 0.075     | $AR(1)$<br>[0.019]    | 1768  | 0.039     |
| Jan 1 1983 to<br>Dec 31 1989 | $AR(0)$<br>[0]            | 1769  | 0.236     | $AR(2)$<br>[0.019]    | 1767  | 0.683     |
| Jan 1 1990 to<br>Dec 31 1996 | $AR(0)$<br>[0]            | 1771  | 0.021     | $AR(2)$<br>[0.003]    | 1769  | 0.063     |
| Jan 1 1997 to<br>Dec 31 2003 | $AR(0)$<br>[0]            | 1761  | 0.035     | $AR(1)$<br>[0.004]    | 1760  | 0.125     |
| Jan 1 2004 to<br>Dec 31 2010 | $AR(2)$<br>[0.125]        | 1763  | 0.498     | $AR(0)$<br>[0]        | 1763  | 0.618     |
| Jan 1 2011 to<br>Dec 31 2017 | $AR(0)$<br>[0]            | 1761  | 0.309     | $AR(0)$<br>[0]        | 1762  | 0.221     |
| Jan 1 2018 to<br>Dec 29 2023 | $AR(0)$<br>[0]            | 1509  | 0.360     | $AR(2)$<br>[0.015]    | 1507  | 0.083     |
| Jan 2 1962 to<br>Dec 29 2023 | $AR(2)$<br>[0.002]        | 15618 | 0.334     | $AR(1)$<br>[.001]     | 12869 | 0.578     |

for

Table 9. Test Results for DUK by Subperiod

|                           | Duke Energy Corporation (DUK) |       |           |
|---------------------------|-------------------------------|-------|-----------|
| Sample Period             | AR(p)<br>[ $R^2$ ]            | T     | Frequency |
| Jan 2 1962 to Dec 31 1968 | AR(1)<br>[0.018]              | 1737  | 0.082     |
| Jan 1 1969 to Dec 31 1975 | AR(0)<br>[0]                  | 1767  | 0.551     |
| Jan 1 1976 to Dec 31 1982 | AR(0)<br>[0]                  | 1769  | 0.185     |
| Jan 1 1983 to Dec 31 1989 | AR(0)<br>[0]                  | 1770  | 0.127     |
| Jan 1 1990 to Dec 31 1996 | AR(0)<br>[0]                  | 1771  | 0.009     |
| Jan 1 1997 to Dec 31 2003 | AR(0)<br>[0]                  | 1761  | 0.232     |
| Jan 1 2004 to Dec 31 2010 | AR(1)<br>[0.024]              | 1762  | 0.128     |
| Jan 1 2011 to Dec 31 2017 | AR(0)<br>[0]                  | 1761  | 0.058     |
| Jan 1 2018 to Dec 29 2023 | AR(0)<br>[0]                  | 1509  | 0.040     |
| Jan 2 1962 to Dec 29 2023 | AR(1)<br>[0.002]              | 15602 | 0.296     |

Table 10. Test Results for VWRETX and EWRETX by Subperiod

|                              | Value Weighted<br>Index (VWRETX) |       |           | Equal Weighted<br>Index (EWRETX) |       |           |
|------------------------------|----------------------------------|-------|-----------|----------------------------------|-------|-----------|
| Sample Period                | $AR(p)$<br>[ $R^2$ ]             | T     | Frequency | $AR(p)$<br>[ $R^2$ ]             | T     | Frequency |
| Jan 2 1962 to<br>Dec 31 1968 | $AR(0)$<br>[0]                   | 1736  | 0.656     | $AR(1)$<br>[0.101]               | 1735  | 0.795     |
| Jan 1 1969 to<br>Dec 31 1975 | $AR(1)$<br>[0.101]               | 1765  | 0.216     | $AR(1)$<br>[0.218]               | 1765  | 0.479     |
| Jan 1 1976 to<br>Dec 31 1982 | $AR(1)$<br>[0.035]               | 1768  | 0.143     | $AR(0)$<br>[0]                   | 1769  | 0.644     |
| Jan 1 1983 to<br>Dec 31 1989 | $AR(1)$<br>[0.014]               | 1768  | 0.024     | $AR(0)$<br>[0]                   | 1769  | 0.665     |
| Jan 1 1990 to<br>Dec 31 1996 | $AR(0)$<br>[0]                   | 1771  | 0.389     | $AR(0)$<br>[0]                   | 1771  | 0.994     |
| Jan 1 1997 to<br>Dec 31 2003 | $AR(0)$<br>[0]                   | 1761  | 0.351     | $AR(0)$<br>[0]                   | 1761  | 0.618     |
| Jan 1 2004 to<br>Dec 31 2010 | $AR(0)$<br>[0]                   | 1763  | 0.450     | $AR(0)$<br>[0]                   | 1763  | 0.791     |
| Jan 1 2011 to<br>Dec 31 2017 | $AR(0)$<br>[0]                   | 1761  | 0.044     | $AR(0)$<br>[0]                   | 1761  | 0.212     |
| Jan 2 2018 to<br>Dec 29 2023 | $AR(0)$<br>[0]                   | 1509  | 0.233     | $AR(0)$<br>[0]                   | 1509  | 0.468     |
| Jan 2 1962 to<br>Dec 29 2023 | $AR(0)$<br>[0]                   | 15605 | 0.261     | $AR(0)$<br>[0]                   | 15605 | 0.604     |