



Tests for m -dependence based on sample splitting methods[☆]

Seongman Moon^{*}, Carlos Velasco¹

Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe Madrid, Spain

ARTICLE INFO

Article history:

Received 24 November 2010

Received in revised form

3 July 2012

Accepted 4 November 2012

Available online 28 December 2012

JEL classification:

C14

F31

F37

Keywords:

m -dependence

Sample splitting

Pooled method

Wald method

Minimum/maximum/median method

Expectations hypothesis

ABSTRACT

This paper develops new test methods for m -dependent data. Our approach is based on sample splitting by regular sampling of the original data at lower frequencies, so that standard techniques for testing independence can be used for each individual subsample. We then propose several alternative statistics that aggregate information across subsamples and investigate their asymptotic and finite sample properties. We apply our methods to test the predictability of excess returns in foreign exchange markets. We also illustrate how our serial dependence tests can provide useful information for identifying particular economic alternatives when testing the expectations hypothesis in foreign exchange markets.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In many contexts it is known or assumed that economic time series are m -dependent. We say that a stochastic process $\{X_t\}_{t=-\infty}^{\infty}$ is m -dependent if for some integer $m \geq 0$ and every n , the collections

$$\{\dots, X_{n-1}, X_n\} \quad \text{and} \quad \{X_{n+m+1}, X_{n+m+2}, \dots\}$$

are independent. In such cases, usual independence is equivalent to 0-dependence. m -dependence arises when data are sampled more finely than the forecasting interval (or maturity) for testing the expectations or the efficient market hypotheses, when differencing methods are applied to remove fixed effects, or when evaluating the goodness-of-fit of a moving average (MA) model.

Our paper is inspired by testing the predictability of k -month ahead excess returns, which is a key step when investigating the expectations hypotheses of forward exchange rates or of the term

structure of interest rates.² If data are sampled more finely, for example monthly, than the forecasting horizon, the forecasting errors then display a MA structure and become m -dependent (or $(k-1)$ -dependent, in this example) under the expectations hypothesis.³ To take into account this dependency, Hansen and Hodrick (1980) examined restrictions on a k -step ahead forecasting regression and proposed corrected standard errors (see also Newey and West (1987)). Researchers, however, have faced difficulties in applying their methods to tests that are designed only for independent data. For instance, Campbell and Dufour (1995) had to assume that forecasting errors are independent to develop tests of orthogonality based on signs and signed ranks. Escanciano and Velasco (2006) who used nonlinear test methods also had to assume unpredictable forecasting errors. The conditional test by Jansson and Moreira (2006) and the Q -test by Campbell and Yogo (2006) are other examples of this issue.

[☆] We are grateful to Associate Editor and two anonymous referees for helpful comments and suggestions.

^{*} Corresponding author. Tel.: +34 91 624 8668; fax: +34 91 624 9329.

E-mail addresses: smoon@eco.uc3m.es (S. Moon), Carlos.Velasco@uc3m.es (C. Velasco).

¹ Tel.: +34 91 624 9646; fax: +34 91 624 9329.

² See Engel (1996) and Lewis (1995) for a survey of tests for the expectations hypothesis in foreign exchange markets; and see Campbell and Shiller (1991) and Cochrane and Piazzesi (2005) for tests of the expectations hypothesis of the term structure of interest rates.

³ We use the notation of m -dependence for the general concept and $(k-1)$ -dependence for a specific example throughout the paper. The notation of m -dependence is from the statistical literature and the notation of k is from the expectations hypothesis literature.

In this paper we develop new testing methods for m -dependent data. Our approach first splits the original time series into $m + 1$ subsamples so that the observations within each subsample are independent under the null hypothesis of m -dependence. Then, we apply standard techniques for testing uncorrelation to the individual subsamples. However, m -dependence induces correlation between the subsamples that must be accounted for when constructing test statistics to aggregate information from the subsamples.

To apply our sample splitting methods, we focus on linear methods based on the autocorrelations most stressed in practice.⁴ In particular, we apply them to three common serial dependence tests: the variance ratio, the Box–Pierce portmanteau, and the Fama and French (1988) tests. Specifically, we first propose Wald type statistics exploiting joint distributions across subsamples. Second, we use Bonferroni bounds to control the size of the tests based on one-sided maximal deviations. Third, we design new statistics that pool estimates from individual subsamples. To improve the finite sample behavior of these test procedures, we design a parametric bootstrap method that accounts for the effects of the m -dependence and show that it provides very good size accuracy even with long return horizons, for which the asymptotic normal approximation usually performs poorly.

Our sample splitting methods have several advantages for addressing m -dependent data compared to previous studies. The residual-based methods using parameter estimates can be subject to estimation errors. For example, for testing a linear MA (m) model, the asymptotic distribution of goodness-of-fit statistics based on residuals autocorrelation might depend on the estimation method employed.⁵ Regression-based methods also need to account for this serial dependence through different variants of autocorrelation robust standard errors. In contrast, our sample splitting methods guarantee exact independence of the data under the null hypothesis of m -dependence because they do not estimate (or even specify) parametric models. Dufour and Torres (1998) proposed sample splitting in the context of regression-based tests, but used bound methods to conduct asymptotic inference, while we explicitly account for the dependence among subsamples avoiding possible efficiency losses due to these approximations.

We also distinguish our methods from the typical long-horizon tests based on variance ratios for the random walk hypothesis of asset prices.⁶ These tests mainly differ from ours in terms of the assumption of m -dependent data. The random walk assumption always implies 0-dependence, regardless of how finely the data are sampled. However, maturities (or forecasting intervals) in the expectations hypothesis induce m -dependence components in excess returns. In this case, the expected value of the usual variance ratio statistic is no longer 1, and is left unspecified. Therefore, direct application of this and related predictability tests requires adjusting the sampling time interval to maturity to guarantee uncorrelated returns under the expectations hypothesis. Consequently, this approach leads to inefficiencies due to an effectively reduced sample size and poses a new problem of aggregating information if several subsamples are used instead.

The rest of the paper is organized as follows. Section 2 presents the asymptotic theory of our sample splitting methods and Section 3 provides the parametric bootstrap procedures. We

employ the sample splitting methods to tests for the expectations hypothesis in foreign exchange markets in Section 4 and discuss their size and power properties in Section 5. An empirical study for testing the expectations hypothesis in foreign exchange markets is provided in Section 6 and concluding remarks follow.

2. The asymptotic theory of sample splitting methods for m -dependent data

In this section, we present sample splitting methods for m -dependent data in the context of testing the expectations hypothesis. We first characterize the notion of the lack of linear predictability beyond a forecasting horizon k and introduce further assumptions that lead to exact $(k - 1)$ -dependence. We then provide the asymptotic properties of the serial dependence tests implemented in the presence of m -dependent data. We begin with the typical variance ratio test. To compare our methods with previous studies, we briefly state the distributional results of the variance ratios when $k = 1$. Here, $k = 1$ means that the variance test assumes 0-dependent data under the null hypothesis. Then, we analyze the asymptotic properties of the variance ratio tests when $k > 1$. Furthermore, we show that the t statistics of Fama and French (1988) belong to the class of generalized variance ratios, and we analyze their asymptotic properties in this context. We also apply our methods to the Box and Pierce (1970) portmanteau Q statistics. Finally, we relate our results to previous studies and discuss the applications of our methods to other tests, possibly capturing nonlinear dependence.

Let $S = \{\xi_{1|k}, \xi_{2|k}, \dots, \xi_{T|k}\}$ be the original sample, where $\xi_{t|k}$ denotes the k -period excess return between period $t - k$ and t . Suppose a researcher wants to test the predictability of excess returns beyond horizon k . Then, the data become $(k - 1)$ -dependent under the null hypothesis, and the usual methods for testing uncorrelation or independence cannot be used directly. For example, if data are collected at a monthly frequency and the researcher wants to forecast excess returns for a holding period of three-months, then $k = 3$. To address this dependence, we propose sample splitting methods. The main idea of our procedure is to first divide the original sample into k subsamples in the following way. Define each of the k subsamples by $S_1 = \{\xi_{1|k}, \xi_{k+1|k}, \dots, \xi_{T-k+1|k}\}$, $S_2 = \{\xi_{2|k}, \xi_{k+2|k}, \dots, \xi_{T-k+2|k}\}, \dots, S_k = \{\xi_{k|k}, \xi_{2k|k}, \dots, \xi_{T|k}\}$. Here, a subsample is constructed such that all $\xi_{t|k}$ are uncorrelated or unpredictable within the subsample under the null of no predictability but a subsample itself is correlated with other subsamples. Then we use the usual tests for uncorrelation or independence for each subsample. Finally, we propose several methods that aggregate information across the subsample statistics.

2.1. An econometric framework

Assume that the k -period excess returns $\xi_{t|k}$ are covariance stationary and have autocorrelation sequence $\gamma_k(i) = \text{Cov}(\xi_{t|k}, \xi_{t+i|k}) / \text{Var}(\xi_{t|k})$ satisfying $\gamma_k(i) = 0$ for $|i| \geq k$. Then, from the Wold decomposition, it holds that

$$H_0^{(k)} : \xi_{t+k|k} = \alpha_k + \sum_{i=1}^k c_i e_{t+i},$$

where e_t is weak noise, i.e., $E[e_t] = 0$, $E[e_t^2] = \sigma^2$ and $E[e_t e_{t-i}] = 0$ for any $i \neq 0$, and $\sigma_k^2 = \text{Var}(\xi_{t|k}) = \sigma^2 \sum_{i=1}^k c_i^2 > 0$. Thus $\xi_{t|k}$ follows a weak linear MA($k - 1$) model. Here, the hypothesis implies that information in or prior to $\xi_{t|k}$ is not useful to forecast $\xi_{t+k|k}$ linearly and that the autocorrelation function is truncated. However, the hypothesis does not restrict the possibility of nonlinear relationships in higher order moments at any horizon.

⁴ Our approach can also permit different characterizations of the independence hypothesis, including the martingale difference and the white noise hypotheses.

⁵ Those statistics include the Box–Pierce, the LM, and the T_p statistics. See Delgado and Velasco (2011) for a recent discussion.

⁶ See, for example, Lo and MacKinlay (1988) and Poterba and Summers (1988) for stock prices, Liu and He (1991) for spot exchange rates, and Cochrane (1988) for the U.S. output.

So e_t still can be nonlinearly predictable at any horizon or can display conditional dynamic heteroscedasticity.

To build asymptotic theory on sample autocorrelations and variance ratios, we need to further restrict the dependence of the innovation process e_t . For this, we impose mixing conditions plus one assumption on the higher joint moments restricting the form of a possible ARCH structure, using Assumption H* of Lo and MacKinlay (1988) provided next.

Assumption 2.1. 1. For all t , $E[e_t] = 0$ and $E[e_t e_{t-\tau}] = 0$ for any $\tau \neq 0$.

2. $\{e_t\}$ is ϕ -mixing with coefficients $\phi(j)$ of size $r/(2r - 1)$ or is α -mixing with coefficients $\alpha(j)$ of size $r/(r - 1)$, where $r > 1$, such that for all t and for any $\tau \geq 0$, there exists some $\delta > 0$ for which $E|e_t e_{t-\tau}|^{2(r+\delta)} < \Delta < \infty$.

3. $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E[e_t^2] = \sigma^2 < \infty$.

4. For all t , any nonzero j and i where $j \neq i$

$$E[e_t^2 e_{t-j} e_{t-i}] = 0. \tag{1}$$

Assumption 2.1 guarantees that the linear projection of $\xi_{t+k|k}$ given $\xi_{t|k}, \xi_{t-1|k}, \dots$ is constant. The mixing conditions in Assumption 2.1.2 can be replaced by a martingale difference assumption stating that $E[e_t | e_{t-1}, e_{t-2}, \dots] = 0$ for all t . This martingale assumption on e_t would imply that the process $\xi_{t|k}$ is not predictable beyond horizon k under $H_0^{(k)}$, i.e., the conditional expectation of $\xi_{t+k|k}$ given $\xi_{t|k}, \xi_{t-1|k}, \dots$ is constant. Note that (1) allows for deterministic changes in the variance and for ARCH effects. In general, this condition implies that the sample autocorrelation coefficients of e_t at different lags are asymptotically uncorrelated, despite the presence of heteroscedasticity. Based on condition (1), Lo and MacKinlay (1988) propose robust estimates of the asymptotic variance of autocorrelation coefficients that lead to asymptotic normal feasible inference for variance ratio tests. If we further impose the homoscedasticity of e_t , the asymptotic variance expressions simplify, as is the case when we strengthen the dependence conditions on e_t to exact independence as in the next assumption.

Assumption 2.2. e_t is an iid random variable with mean zero, variance σ^2 and finite fourth moment.

When Assumption 2.2 holds, the process $\xi_{t|k}$ is exactly $(k - 1)$ -dependent under $H_0^{(k)}$, but it also imposes a linear conditional expectation for $\xi_{t|k}$ up to the forecasting horizon k . Thus, $H_0^{(k)}$ under Assumption 2.2 is strictly stronger than

$\mathcal{H}_0^{(k)} : \{\xi_{t|k}\}$ is a stationary $(k - 1)$ -dependent process.

Nevertheless, any type of (possibly nonlinear) dependence is ruled out beyond horizon k in both cases.

Tests for $H_0^{(k)}$ can be designed using autocorrelations at lags beyond $k - 1$, but their asymptotic properties are affected by the $(k - 1)$ -dependence. In the remainder of this section, we investigate the asymptotic properties of several tests under $H_0^{(k)}$ and $\mathcal{H}_0^{(k)}$, first for $k = 1$, and then for $k > 1$ using sample splitting. In any case, we focus on testing for linear dependence based on the autocorrelations.⁷

2.2. Variance ratio statistics when $k = 1$

The null hypothesis $H_0^{(1)}$ can characterize both the expectations and the random walk hypotheses, referring either to the levels of

increments of a given process. So, the traditional variance ratio tools are useful to measure the deviations from these hypotheses. Define the population variance ratio $VR_1(q)$ of the one-period excess return $\xi_{t|1}$, exploiting the fact that the variance of the sum of q consecutive excess returns should be q times larger than that of $\xi_{t|1}$ under the null hypothesis $H_0^{(1)}$,

$$VR_1(q) = \frac{\text{Var}\left(\sum_{i=0}^{q-1} \xi_{t+i|1}\right)}{q \text{Var}(\xi_{t|1})} = 1 + 2 \sum_{i=1}^{q-1} \left(1 - \frac{i}{q}\right) \gamma_1(i), \tag{2}$$

where q is a positive integer aggregation value and $\gamma_1(i) = \text{Cov}(\xi_{t|1}, \xi_{t+i|1}) / \text{Var}(\xi_{t|1})$ denotes the autocorrelation of excess returns between time t and $t + i$. $VR_1(q)$ should be equal to one as long as the excess returns are not serially correlated. If the returns are positively (negatively) correlated, $VR_1(q)$ should be larger (less) than one.

Now define the corresponding sample variance ratio statistic as

$$\widehat{VR}_1(q) = \frac{\widehat{\sigma}_{b|1}^2(q)}{\widehat{\sigma}_{a|1}^2}, \tag{3}$$

where $\widehat{\sigma}_{b|1}^2(q) = (qg_1(q))^{-1} \sum_{t=q}^T (\xi_{t|1} + \xi_{t-1|1} + \dots + \xi_{t-q+1|1} - q\widehat{\alpha}_1)^2$ and $\widehat{\sigma}_{a|1}^2 = \widehat{\sigma}_{b|1}^2(1)$. Here, T is sample size, $g_1(q) = (T - q + 1)(1 - q/T)$ corrects the bias in the variance estimator $\widehat{\sigma}_{b|1}^2(q)$ under the null, and $\widehat{\alpha}_1 = T^{-1} \sum_{t=1}^T \xi_{t|1}$. Because the mean and variance of the q consecutive returns are linear in the aggregation interval q under the null hypothesis $H_0^{(1)}$, $\widehat{\sigma}_{b|1}^2(q)$ is an unbiased estimate of the variance of a single return. In this sense, the variance ratio test on uncorrelated excess returns shares the essence of the random walk hypothesis test, which exploits that the variance of random walk increments must be a linear function of the time interval.

Based on Lo and MacKinlay's (1988) analysis, a variance ratio test for the expectations hypothesis can be easily developed so that

$$z_1(q) = \sqrt{T} (\widehat{VR}_1(q) - 1) \left(\frac{2(2q - 1)(q - 1)}{3q} \right)^{-\frac{1}{2}}$$

follows the standard normal distribution asymptotically under $H_0^{(1)}$ and Assumption 2.2, and further t statistics can be developed under Assumption 2.1.

Note that economic theories for both hypotheses do not explicitly guide the choice of q . In the long-horizon predictability tests for the random walk of asset prices, the choice of large values of q is motivated by the desire to detect the effect of a highly persistent component in asset prices on the return predictability and thus to improve the power of these tests. However, in principle, $q = 2$ can be enough to test the expectations hypothesis, but examining the serial dependence pattern across different q can provide useful information for the identification of an alternative hypothesis as discussed in Section 4. Regardless of the objectives of the study, the choice of q in both tests involves the use of overlapping observations as is explicit in the definition of $\widehat{\sigma}_{b|1}^2(q)$. However, as shown in detail below, the nature of the dependence arising in this context is different from that induced by subsampling when $k > 1$. One of our objectives in this paper is to address this difference.

2.3. Variance ratio statistics when $k > 1$

The above approach can only be used for the case in which the maturity or the forecasting interval exactly matches the sampling time interval. However, it is not uncommon for researchers to use data that are sampled more finely than the forecasting interval or maturity for testing the predictability of excess returns. In this section, we assume $k > 1$. That is, we are concerned with the case for testing the lack of linear predictability beyond a forecasting horizon k .

⁷ This linear approach may lose consistency against some nonlinear dependence alternatives. But, many general tests focusing on martingale difference conditions could be adapted to the present context similar to the way we describe for correlation-based statistics [see, for example, Escanciano and Velasco (2006)].

To address this m -dependent data, as shown in the beginning of this section, the original sample is first split into k subsamples. The traditional variance ratio test is then used for each subsample because they contain uncorrelated observations under the null. We now describe several methods to aggregate the information contained in all of the subsamples: the Wald method, tests based on extreme values and Bonferroni bounds, and pooled tests. Most of the methods that we propose have standard asymptotic distributions under the null, but could have different behavior under general alternatives. These variations would lead to specific recommendations in favor of some particular methods over others in applied work. To develop these recommendations we conduct extensive Monte Carlo experiments in Section 5.

We begin with a general result for the variance estimates constructed from individual subsamples, where

$$\widehat{V}_{j|k}(q) = \frac{1}{qg_k(q)} \sum_{t=q}^{T/k} (\xi_{k(t-1)+j|k} + \xi_{k(t-2)+j|k} + \dots + \xi_{k(t-q)+j|k} - q\widehat{\alpha}_{j|k})^2$$

and $\widehat{\alpha}_{j|k} = (k/T) \sum_{t=1}^{T/k} \xi_{k(t-1)+j|k}$ depends only on subsample $j, j = 1, \dots, k$. The factor $g_k(q) = ((T/k) - q + 1)(1 - q/(T/k))$ corrects the biases in the variance estimator $\widehat{V}_{j|k}(q)$ caused by both the use of overlapping q -period excess returns and the mismatch between forecasting and sampling intervals.⁸ The unbiasedness can be easily checked because we construct $\widehat{V}_{j|k}(q)$ from subsample j , which contains only uncorrelated observations under the null hypothesis in an analogous way to Eq. (3).

Lemma 2.1. Under $\mathcal{H}_0^{(k)}$, $\widehat{V}_{j|k}(q), j = 1, \dots, k$ are consistent and unbiased for σ_k^2 for each positive integer q .

Note that it is possible to test the null hypothesis by only employing information in a given subsample, while dropping the other observations, using the individual variance ratio statistics:

$$\widehat{VR}_k^{(j)}(q) = \frac{\widehat{V}_{j|k}(q)}{\widehat{V}_{j|k}(1)}, \quad j = 1, \dots, k.$$

However, a single subsample contains only T/k observations, which is a fraction of the original sample. Instead, our modified approaches shown below increase the effective sample size by k times and can yield important efficiency gains. For example, k is 3 when monthly observations are used for testing the three-month excess return predictability but k becomes 13 when weekly observations are used.

To exploit simultaneously all $\widehat{VR}_k^{(j)}(q)$ available in a given data set, we consider the asymptotic joint distribution of

$$U_k(q) = \frac{\sqrt{T/k}}{\sqrt{2(q-1)(2q-1)/3q}} \times \left(\widehat{VR}_k^{(1)}(q) - 1, \dots, \widehat{VR}_k^{(k)}(q) - 1 \right)'. \quad (4)$$

Denote as $\delta_k^{(a,b)}(i, j)$ the asymptotic covariance of sample autocorrelations at lags i and j across the different subsamples a and b ,

$$\delta_k^{(a,b)}(i, j) = ACov \left((T/k)^{1/2} \widehat{\gamma}_k^{(a)}(i), (T/k)^{1/2} \widehat{\gamma}_k^{(b)}(j) \right).$$

Lemma 2.2. Under $\mathcal{H}_0^{(k)}$,

$$U_k(q) \sim_a N(0, \Sigma_k(q)),$$

where the diagonal elements of $\Sigma_k(q) > 0$ are 1, and in general, $1 \leq b \leq a \leq k$,

$$\Sigma_k(q)^{[a,b]} = \frac{1}{\sigma_k^4} \left\{ \sum_{i=1}^{q-1} \left(1 - \frac{i}{q}\right)^2 \delta_k^{(a,b)}(i, i) + \sum_{i=2}^{q-1} \left(1 - \frac{i}{q}\right) \left(1 - \frac{i-1}{q}\right) \times \left[\delta_k^{(a,b)}(i, i-1) + \delta_k^{(a,b)}(i-1, i) \right] \right\},$$

with $\sigma_k^2 = Var(\xi_{t|k})$ and

$$\begin{aligned} \delta_k^{(a,b)}(i, i) &= E \left[\bar{\xi}_0 \bar{\xi}_{a-b} \bar{\xi}_{ik} \bar{\xi}_{ik+a-b} \right] + E \left[\bar{\xi}_0 \bar{\xi}_{a-b-k} \bar{\xi}_{ik} \bar{\xi}_{ik+a-b-k} \right], \quad i > 0; \\ \delta_k^{(a,b)}(i, i-1) &= E \left[\bar{\xi}_0 \bar{\xi}_{a-b} \bar{\xi}_{ik} \bar{\xi}_{ik+a-b-k} \right], \quad i > 1; \\ \delta_k^{(a,b)}(i, i+1) &= E \left[\bar{\xi}_0 \bar{\xi}_{a-b-k} \bar{\xi}_{ik} \bar{\xi}_{ik+a-b} \right], \quad i > 0; \end{aligned} \quad (5)$$

for $\bar{\xi}_t = \xi_{t|k} - \alpha_k$, and $\delta_k^{(a,a)}(i, j) = 0, i \neq j; \delta_k^{(a,b)}(i, j) = 0, |i-j| > 1$.

The correlation among different subsamples is reflected in the terms depending on $\delta_k^{(a,b)}(i, j)$ for $a \neq b$ in $\Sigma_k(q)$. However, if $a = b, \delta_k^{(a,a)}(i, i) = E \left[\bar{\xi}_0^2 \right] E \left[\bar{\xi}_{ik}^2 \right] = E \left[\bar{\xi}_0^2 \right]^2 = \sigma_k^4$, but $\delta_k^{(a,a)}(i, j) = 0$ for $i \neq j$ under $\mathcal{H}_0^{(k)}$ so that $\Sigma_k(q)^{[a,a]} = 1$, recovering the usual result under independence, $k = 1$.

In general, for any a, b , and $i > 1$, the first element in $\delta_k^{(a,b)}(i, i)$ factorizes under $\mathcal{H}_0^{(k)}$, i.e., $E \left[\bar{\xi}_0 \bar{\xi}_{a-b} \bar{\xi}_{ik} \bar{\xi}_{ik+a-b} \right] = E \left[\bar{\xi}_0 \bar{\xi}_{a-b} \right] E \left[\bar{\xi}_{ik} \bar{\xi}_{ik+a-b} \right] = E \left[\bar{\xi}_0 \bar{\xi}_{a-b} \right]^2$, but does not factorize when $i = 1$ because it is not possible to isolate pairs of independent random variables in $E \left[\bar{\xi}_0 \bar{\xi}_{a-b} \bar{\xi}_{ik} \bar{\xi}_{ik+a-b} \right]$. However, all other expectations in $\Sigma_k(q)$ factorize for all values of i indicated:

$$\begin{aligned} E \left[\bar{\xi}_0 \bar{\xi}_{a-b-k} \bar{\xi}_{ik} \bar{\xi}_{ik+a-b-k} \right] &= E \left[\bar{\xi}_0 \bar{\xi}_{a-b-k} \right] E \left[\bar{\xi}_0 \bar{\xi}_{k-a+b} \right] \\ &= E \left[\bar{\xi}_0 \bar{\xi}_{a-b-k} \right]^2, \quad i > 0; \\ \delta_k^{(a,b)}(i, i-1) &= E \left[\bar{\xi}_0 \bar{\xi}_{a-b} \right] E \left[\bar{\xi}_0 \bar{\xi}_{a-b-k} \right], \quad i > 1; \\ \delta_k^{(a,b)}(i, i+1) &= E \left[\bar{\xi}_0 \bar{\xi}_{a-b-k} \right] E \left[\bar{\xi}_0 \bar{\xi}_{a-b} \right], \quad i > 0, \end{aligned}$$

so that the right hand side of the above three terms does not depend on i , and the estimation of $\Sigma_k(q)$ is simplified under $\mathcal{H}_0^{(k)}$.

If we impose linearity, but allow for general dynamic heteroscedasticity, the basic results on the asymptotic distribution of U_k hold, but $\Sigma_k(q)$ is affected. In particular, under $H_0^{(k)}$, condition (1) implies that $E \left[\bar{\xi}_t^2 \bar{\xi}_{t-j} \bar{\xi}_{t-i} \right] = 0$ for $|j-i| \geq k, j > k$, and $i > k$, leading to the following corollary.

Corollary 2.3. Under $H_0^{(k)}$ and Assumption 2.1, the conclusions of Lemma 2.2 hold, but the diagonal elements of $\Sigma_k(q)$ are not necessarily equal to one.

The main difference with respect to the case of exact m -dependence in Lemma 2.2 is that all of the expectations in the terms $\delta_k^{(a,b)}(i, j)$ depend now on i and j and on which particular subsamples a and b are involved in. Therefore, no general factorizations are possible in this case due to the possible presence of conditional heteroscedasticity effects. For instance, even if $a = b, E \left[\bar{\xi}_0 \bar{\xi}_{a-b} \bar{\xi}_{ik} \bar{\xi}_{ik+a-b} \right] = E \left[\bar{\xi}_0^2 \bar{\xi}_{ik}^2 \right] \neq \sigma_k^4$, because of possible correlation between $\bar{\xi}_0^2$ and $\bar{\xi}_{ik}^2$ (for all $i > 0$), so that $\Sigma_k(q)$ has no longer elements equal to one in the main diagonal.

However, if we impose both linearity and conditional homoscedasticity then the following corollary follows, further exploiting these factorization properties.

⁸ For notational simplicity we assume that T/k is integer.

Corollary 2.4. Under $H_0^{(k)}$ and Assumption 2.2, the conclusions of Lemma 2.2 hold with, $1 \leq b \leq a \leq k$,

$$\Sigma_k(q)^{[a,b]} = \frac{1}{\sigma_k^4} \left\{ E[\bar{\xi}_0 \bar{\xi}_{a-b}]^2 + E[\bar{\xi}_0 \bar{\xi}_{k-a+b}]^2 + \frac{4(q-2)}{2q-1} E[\bar{\xi}_0 \bar{\xi}_{a-b}] E[\bar{\xi}_0 \bar{\xi}_{k-a+b}] \right\}, \quad (6)$$

where $E[\bar{\xi}_0 \bar{\xi}_s] = \sigma^2 \sum_{i=1}^{k-s} c_i c_{i-s}$, $0 \leq s < k$, and $E[\bar{\xi}_0 \bar{\xi}_k] = 0$.

Lemma 2.2 and its corollaries provide many alternative ways to devise tests using the entire sample of size T . One approach is to use a Wald type statistic, as proposed by Richardson and Smith (1991) in a related context. The class of Wald statistics is defined by

$$W_k(q; R) = (RU_k(q))' (R\widehat{\Sigma}_k(q)R')^{-1} RU_k(q), \quad (7)$$

where R is a full row-rank non-random $r \times k$ matrix. $W_k(q; R)$ is asymptotically distributed as a χ_k^2 variable under the null for $\widehat{\Sigma}_k(q) \rightarrow_p \Sigma_k(q)$.⁹ Consistent estimates $\widehat{\Sigma}_k(q)$ can be obtained by sample analogs of the expectations in $\delta_k^{(a,b)}$. The standard case is when $R = I_k$, involving tests for the joint hypothesis of all individual variance ratios being equal to one. Taking $R = (1/k, \dots, 1/k)$ we can test whether the average variance ratio across subsamples is equal to one (see the detailed discussion below). Setting $r = 1$, we can also provide t -tests for each individual variance ratio.

A further approach to summarize the information of variance ratios in subsamples can be based on the extreme statistics of $U_k(q)$,

$$\text{Max}z_k(q) = \max U_k(q), \quad \text{Min}z_k(q) = \min U_k(q). \quad (8)$$

Using the max and min statistics we can perform one-sided tests, right and left tests, respectively, based on the normal asymptotic critical values with a significance level α/k , invoking Bonferroni inequality. Alternatively, one might wish to further exploit the information on excess returns contained in the distribution of subsample variance ratios, by looking at other summary statistics. Based on the joint distribution of subsamples, for example, either calendar or seasonal effects on excess returns can also be examined.

Finally, we describe a pooled variance ratio statistic using the estimates $\widehat{\sigma}_{b|k}^2(1)$ and $\widehat{\sigma}_{b|k}^2(q)$,

$$\widehat{\sigma}_{b|k}^2(1) = \frac{1}{k} \sum_{j=1}^k \widehat{V}_{jk}(1) \quad \text{and} \quad \widehat{\sigma}_{b|k}^2(q) = \frac{1}{k} \sum_{j=1}^k \widehat{V}_{jk}(q),$$

where $\widehat{\sigma}_{b|k}^2(q)$ are consistent and unbiased for σ_k^2 under the null hypothesis. Then, the pooled sample variance ratio of the k -period excess returns is given by

$$\widehat{VR}_k(q) = \frac{\widehat{\sigma}_{b|k}^2(q)}{\widehat{\sigma}_{b|k}^2(1)}, \quad (9)$$

whose asymptotic distribution is described in the next lemma.

⁹ Richardson and Smith constructed the covariance matrix of variance ratios for various values of q when $k = 1$. In contrast, we construct the covariance matrix of variance ratios in subsamples for a given q . Their aim was to improve the efficiency of the tests based on predictive regressions by using aggregated observations, which will induce MA errors in the transformed regressions. In that framework, the original prediction errors are uncorrelated, so the variance-covariance matrix of the OLS slope estimates with a different number of aggregated observations does not depend on further unknown parameters. However, this does not need to be the case in the presence of a $MA(k-1)$ structure due to the mismatch between maturities and sampling time intervals.

Lemma 2.5. Under $\mathcal{H}_0^{(k)}$,

$$\frac{\sqrt{T}(\widehat{VR}_k(q) - 1)}{\sqrt{2(q-1)(2q-1)/3q}} \sim_a N(0, \Lambda_k(q))$$

where $\Lambda_k(q) > 0$ with

$$\Lambda_k(q) = \frac{6q}{\sigma_k^4(q-1)(2q-1)} \sum_{a=1}^k \sum_{b=1}^k \sum_{i=1}^q \sum_{j=1 \vee i-1}^{q \wedge i+1} \left(1 - \frac{i}{q}\right) \times \left(1 - \frac{j}{q}\right) \delta_k^{(a,b)}(i, j),$$

for $\delta_k^{(a,b)}(i, j)$ given in Eq. (5). Then,

$$z_k(q) = \Lambda_k(q)^{-\frac{1}{2}} \frac{\sqrt{T}(\widehat{VR}_k(q) - 1)}{\sqrt{2(q-1)(2q-1)/3q}} \quad (10)$$

follows the standard normal distribution asymptotically.

From Lemma 2.5, the following corollary on the asymptotic distribution of the pooled variance ratio of k -period excess returns under linearity is an immediate consequence.

Corollary 2.6. Under $H_0^{(k)}$ and Assumption 2.1, the conclusions of Lemma 2.5 hold.

The expression for $\Lambda_k(q)$ simplifies under linearity and conditional homoscedasticity as given in the next corollary.

Corollary 2.7. Under $H_0^{(k)}$ and Assumption 2.2, $\Lambda_k(q) = 1 + \Omega_k(q) > 0$, where

$$\Omega_k(q) = \frac{2}{\sigma_k^4} \sum_{i=1}^{k-1} \frac{k-i}{k} \left\{ E[\bar{\xi}_0 \bar{\xi}_i]^2 + E[\bar{\xi}_0 \bar{\xi}_{k-i}]^2 + \frac{4(q-2)}{(2q-1)} E[\bar{\xi}_0 \bar{\xi}_i] E[\bar{\xi}_0 \bar{\xi}_{k-i}] \right\}.$$

The term $\Omega_k(q)$ appears only due to the correlation between the different k subsamples. Obviously, this term appears neither in the asymptotic distribution of the individual variance ratios nor in the tests for the random walk hypothesis that are used in the typical long-horizon predictability tests.

Finally, it is easy to show that the average Wald statistic $W_k(q; R_0)$ with $R_0 = (1/k, \dots, 1/k)$ in (7) is asymptotically equivalent to the square of the $z_k(q)$ pooled statistic in (10). Note that $\Lambda_k(q) = (1/k) \sum_{a=1}^k \sum_{b=1}^k \Sigma_k(q)^{[a,b]}$ and $\sqrt{T}(\widehat{VR}_k(q) - 1) = \sqrt{T}(\widehat{\sigma}_{b|k}^2(q) - \widehat{\sigma}_{b|k}^2(1)) / \widehat{\sigma}_{b|k}^2(1) = \sqrt{T}(\widehat{\sigma}_{b|k}^2(q) - \widehat{\sigma}_{b|k}^2(1)) / \sigma_k^2 + o_p(1)$. A similar expression holds for $R_0 U_k(q)$, up to scale, so that only the standardization changes between both statistics.

2.4. The Fama–French test and generalized variance ratios

We have considered variance ratio statistics for testing serial dependence. We now present two other serial dependence test statistics, the Fama and French (1988) t -statistics and the Box and Pierce (1970) Q statistics, and show how to implement them in the presence of m -dependent data. We discuss the asymptotic properties of the Fama and French (1988) t -statistics in this subsection and those of the Box–Pierce statistics in the next subsection.

Fama and French (1988) regress the n -period future returns on the n -period past returns to capture a slowly mean reverting component in stock prices:

$$\tilde{\xi}_{t+n,k}^{n,k} = \alpha_{n,k} + \beta_{n,k} \tilde{\xi}_t^{n,k} + u_{t+n,k}^{n,k}, \quad \text{for positive integer } n, \quad (11)$$

where $\tilde{\xi}_t^{n,k} = \sum_{i=0}^{n-1} \xi_{t-ik|k}$ denotes the stock returns between t and $t - (n - 1)k$. The implication of the null hypothesis tested

here is $\beta_{n,k} = 0$ for each n . The Fama–French test was designed to test the random walk hypothesis of stock prices, i.e., $k = 1$ is assumed in the regression. However, this test can also be used for m -dependent data, with a modification that takes into account the mismatch between the forecasting horizon interval and the sampling interval. One way to implement the test for m -dependent data is to run the OLS regression while the standard errors of the slopes are adjusted for the $(nk - 1)$ autocorrelations in the residuals using the method of either Hansen and Hodrick (1980) or Newey and West (1987).

Note that the slope coefficient for the n -period future returns from this Fama–French regression can be transformed into a particular variance ratio deviation if the length of the base period changes. This fact can be easily shown by rewriting the definition of the least squares estimate in Eq. (11) using the population variance ratio deviation of the one-period excess return, $VR_1(q = 2n, q' = n)$, analogous to Eq. (2), while taking into account the change in the base period and assuming $k = 1$,

$$\beta_{n,1} = \frac{Cov(\tilde{\xi}_t^{n,1}, \tilde{\xi}_{t+n}^{n,1})}{Var(\tilde{\xi}_t^{n,1})} = \frac{Var(\tilde{\xi}_t^{n,1} + \tilde{\xi}_{t+n}^{n,1})}{2Var(\tilde{\xi}_t^{n,1})} - 1 = VR_1(q = 2n, q' = n) - 1, \tag{12}$$

where we define the class of generalized variance ratios by

$$VR_1(q, q') = \frac{Var\left(\sum_{i=0}^{q-1} \xi_{t+i|1}\right) / q}{Var\left(\sum_{i=0}^{q'-1} \xi_{t+i|1}\right) / q'} \tag{13}$$

In Eq. (13) q' is the length of the base period, q is the aggregation value, and $VR_1(q, q')$ is defined such that it is equal to one if the excess returns are not serially correlated. When $q' = 1$, we obtain the usual variance ratio.

Then, analogous to Eq. (3), the corresponding sample variance ratio can be defined by

$$\widehat{VR}_1(q, q') = \frac{\widehat{\sigma}_{b_{11}}^2(q)}{\widehat{\sigma}_{b_{11}}^2(q')},$$

which leads to asymptotically equivalent tests of the Fama–French regression, with potential differences in the calculation of standard errors. Finally, we summarize the basic asymptotic properties of the generalized ratio statistics for $k = 1$ in the next result, which is a direct extension of Lo and MacKinlay's (1988) results.

Lemma 2.8. Under $H_0^{(1)}$ and Assumption 2.2,

$$\sqrt{T}(\widehat{VR}_1(q, q') - 1) \sim_a N\left(0, \frac{2(q - q')(2qq' - 2q^2 + 1)}{3qq'}\right).$$

When $q = 2q'$ and $q' = n$, the asymptotic variance of $\widehat{VR}_1(q, q')$ is $(1 + 2n^2)/3n$, which increases with n . One important issue is whether there is any potential advantage to choosing $q' > 1$ in terms of power. The problem is that, for a given q , $\widehat{\sigma}_{b_{11}}^2(q')$ can also change under the alternative. For example, suppose that the excess returns only have a nonzero first autocorrelation. Then, both $\widehat{\sigma}_{b_{11}}^2(q')$ and $\widehat{\sigma}_{b_{11}}^2(q)$ incorporate this correlation for all $q > q' > 1$, i.e.,

$$\widehat{\sigma}_{b_{11}}^2(q) \rightarrow_p Var(\xi_{t+i|k}) + \frac{2(q - 1)}{q} Cov(\xi_{t|k}, \xi_{t+1|k}),$$

so that

$$\widehat{VR}_1(q, q') - 1 \rightarrow_p \frac{2(q - q')}{q'q} \gamma_1(1),$$

and the probability limit of the corresponding t -statistic (scaled by $(q/T)^{1/2}$) is

$$\left(\frac{2(q - q')(2qq' - 2q^2 + 1)}{3qq'}\right)^{-1/2} \frac{2(q - q')}{q'} \gamma_1(1),$$

which gets smaller in absolute value as q' increases. For instance, if $q' = 1, q = 2n$, this limit is $((4n - 1) / (6(2n - 1)))^{-1/2} \gamma_1(1)$, which tends to $(1/3)^{-1/2} \gamma_1(1)$ as n increases, while if $q' = n, q = 2n$, the limit is $((2n^2 + 1) / 6)^{-1/2} \gamma_1(1)$, which is smaller for any $n > 1$ and tends to zero as n increases. Our experiments suggest that this analysis still holds for a wide range of cases in which there are nonzero higher order correlations, but we do not pursue a general result further.

The results in Lemma 2.8 can be easily extended to the cases of $k > 1$ following the methods in Section 2.2. We do not follow this line of research given the potential disadvantages of letting $q' > 1$.

2.5. Autocorrelations and the Box–Pierce test

Variance-ratio statistics can be obtained as weighted averages of sample autocorrelations from lag 1 up to lag $q - 1$, as in Eq. (2). The sign of $\widehat{VR}_k(q)$, however, can be unclear when the null hypothesis fails but true autocorrelations have different signs over q . This implies that the autocorrelations of different sign can cancel out when calculating variance ratio statistics over q . To alleviate the problem that arises when there is no predominant sign in the autocorrelation structure of returns, we can use statistics that do not depend on the sign of $\hat{\gamma}_k(i)$. One of the simplest ways to achieve this property is to use the Box and Pierce (1970) portmanteau statistic,

$$Q(q) = T \sum_{i=1}^q \hat{\gamma}(i)^2,$$

or the variant by Ljung and Box (1978), which aggregates squared autocorrelations with changing weights to improve asymptotic χ^2 approximations,

$$L(q) = T \sum_{i=1}^q \frac{T + 2}{T - i} \hat{\gamma}(i)^2.$$

Because the distribution of $Q(q)$ and $L(q)$ can be approximated by a χ_q^2 variable under the null of iid returns, these statistics lead to one-sided tests that are consistent against any deviation from the null which implies nonzero autocorrelations up to lag q . In fact, the Box–Pierce test is asymptotically equivalent to the Lagrange Multiplier test for correlations up to order q in a Gaussian environment [for example, see Godfrey (1978)].

However, in the presence of m -dependence, the Box–Pierce type tests cannot be used directly to test the null hypothesis of unpredictable excess returns. To resolve this issue, we propose the same approach as that used for the variance ratio statistics, namely, we decompose the original sample into k subsamples consisting of uncorrelated data under the null and then explore different aggregation methods based, for example, on the joint distribution of individual subsample Box–Pierce statistics,

$$Q_k^{(a)}(q) = \frac{T}{k} \sum_{i=1}^q \hat{\gamma}_k^{(a)}(i)^2, \quad a = 1, \dots, k, \tag{14}$$

each of them being χ_q^2 asymptotically.

We then devise tests using the maximum of the individual statistics in Eq. (14),

$$MaxQ_k(q) = \max_a Q_k^{(a)}(q). \tag{15}$$

The null asymptotic distribution of each $Q_k^{(a)}(q)$ is still χ_q^2 , but these statistics are no longer independent across subsamples under m -dependence. Tests based on $MaxQ_k(q)$ using Bonferroni adjusted asymptotic critical values of individual $Q_k^{(a)}(q)$ will provide a conservative testing procedure, but bootstrap methods can be easily applied to exploit the joint distribution as discussed in Section 3.

Alternatively, we can build pooled Box–Pierce statistics based on the joint estimation of autocorrelations,

$$Q_k(q) = T \sum_{i=1}^q \hat{\gamma}_k(i)^2,$$

where

$$\hat{\gamma}_k(i) = \frac{1}{k} \sum_{a=1}^k \hat{\gamma}_k^{(a)}(i).$$

Here, the pooled autocorrelations $\hat{\gamma}_k(i)$ are also asymptotically normal, and a modified version of $Q_k(q)$ accounting for the appropriate standardization of all $\hat{\gamma}_k(i)$ is still asymptotically χ_q^2 as described in next lemma.

Lemma 2.9. Under $H_0^{(k)}$ and Assumption 2.1 or $\mathcal{H}_0^{(k)}$, the asymptotic distribution of the modified pooled $\widehat{Q}_k(q)$ statistic is

$$\widehat{Q}_k(q) = T \hat{\gamma}'_k \Xi_k(q)^{-1} \hat{\gamma}_k \sim_a \chi_q^2 \tag{16}$$

where $\hat{\gamma}_k = (\hat{\gamma}_k(1), \dots, \hat{\gamma}_k(q))'$ and the elements of $\Xi_k(q)$ are

$$\Xi_k^{(i,j)}(q) = \frac{1}{k^2} \sum_{a=1}^k \sum_{b=1}^k \delta_k^{(a,b)}(i,j), \quad i, j = 1, \dots, q,$$

with $\delta_k^{(a,b)}(i,j)$ given in Eq. (5).

To apply this statistic to the cases where conditional heteroscedasticity is allowed, $\Xi_k^{(i,j)}$ can be estimated by plugging in the estimates of the asymptotic covariances $\delta_k^{(a,b)}(i,j)$ of the sample autocorrelations.

2.6. Discussion

We have focused on applying our aggregation methods across subsamples to three serial dependence tests. The same idea can also be used for other tests such as regression and general non-parametric tests. For example, we can employ our methods to regression (11) or to regression-based predictability tests by implementing the Wald, maximum/minimum, or pooled methods, based on the different coefficient estimates, such as OLS, Campbell and Yogo (2006) and Jansson and Moreira (2006).

Furthermore, one can adapt these methods to the sign and signed rank test by Campbell and Dufour (1995) under m -dependence in a similar way. Campbell and Dufour proposed conditional independence tests with exact finite sample distribution under the null. These are nonparametric tests based on signs and ranks that replace observed data and residuals, being valid under general forms of non-normality and conditional heteroscedasticity. In the presence of m -dependent disturbances, these tests can only be used directly on subsamples, leading to the usual aggregation problem.

From a related perspective, the variance ratio statistics using the ranks of Wright (2000)¹⁰ can be especially useful in the

presence of data with either outliers or important non-normality features that can affect the precision of the asymptotic results or even their validity if higher order moments are not finite. Let $r_j(\xi_{t|k})$ be the rank of $\xi_{t|k}$ among all elements $\xi_{j|k}, \xi_{k+j|k}, \dots, \xi_{T-k+j|k}$ in subsample j . Then, a simple linear transformation of the ranks $r_j(\xi_{t|k})$ is defined by

$$r_t = \left(r_j(\xi_{t|k}) - \frac{T/k + 1}{2} \right) \left(\frac{(T/k - 1)(T/k + 1)}{12} \right)^{-1/2},$$

where r_t is standardized with sample mean 0 and variance 1. The rank-based variance ratio statistic $\widehat{V}_{jk}^r(q) = (qg_k(q))^{-1} \sum_{t=q}^{T/k} (r_{k(t-1)+j} + \dots + r_{k(t-q)+j})^2$ in subsample j is obtained by simply substituting r_t for $\xi_{t|k}$ in $\widehat{V}_{jk}(q)$. Let

$$U_k^r(q) = \frac{\sqrt{T/k}}{\sqrt{2(q-1)(2q-1)/3q}} (\widehat{V}_{1|k}^r(q) - 1, \dots, \widehat{V}_{k|k}^r(q) - 1)'$$

The denominator $\widehat{V}_{jk}^r(1)$, corresponding to $\widehat{V}_{jk}(1)$ in $U_k(q)$ in Eq. (4), is omitted because it is equal to 1 by construction. Then, the rank-based maximum, minimum, and median variance ratios are calculated by

$$\begin{aligned} Maxz_k^r(q) &= \max U_k^r(q), \quad Minz_k^r(q) = \min U_k^r(q), \\ Medz_k^r(q) &= \text{median } U_k^r(q). \end{aligned} \tag{17}$$

We now relate our results to the long-horizon tests based on the variance ratios. It is now well established that the finite-sample distribution of variance ratios and autocorrelation statistics can be quite different from the usual asymptotic approximations due to overlap in the returns data, in particular with a small number of non-overlapping asset returns. For example, Richardson and Stock (1989) show that sample variance ratios are not consistent if q/T approaches some constant and that asymptotic results based on fixed q theory perform poorly in finite samples. Our sample splitting methods would not solve this problem and should be limited to cases in which $q/(T/k)$ is reasonably small. To avoid any confusion, we emphasize that the efficiency gains from our methods are achieved by exploiting all subsamples for a given value of $q/(T/k)$, which can work well in testing the expectations hypothesis with relatively small values of q .¹¹ If there is concern regarding the poor finite sample properties of the conventional variance ratios bootstrap methods can alternatively be used to alleviate those size distortions in finite samples. We show that a parametric bootstrap method presented in the next section obtains quite reasonable size properties even for longer return horizons in a variety of situations.

We finally discuss some possible power disadvantages of our sample splitting methods. If higher order dependence only occurs at some specific lags, e.g., over lag $k + 1$, the tests based on our methods might not detect this dependence. One solution to this identification problem is to use sampling at longer intervals such as $k + 1, k + 2$, etc., which renders subsamples with independent observations under the null and spans wider ranges of dependence. These alternative sample splitting schemes still generate dependence between subsamples, which could be accounted for using similar methods.

¹⁰ Wright provides several alternative variance ratio tests using the ranks and signs of a time series. In principle, our methods can be applied to all of his tests. For the sake of simplicity, we only discuss the application of rank-based variance ratio tests.

¹¹ Richardson and Stock (1989) assume that asset prices follow a random walk so that the analysis is limited to $k = 1$. In this case, increasing T using higher frequency data would not solve the issue because q will also increase proportionally. See, for example, Campbell et al. (1997, p. 79) for the discussion on this matter. On the other hand, our focus is on cases of $k > 1$, which is dictated, e.g., by maturity (but not a research choice). So, the use of higher frequency data can obtain power gains by exploiting information from increasing the number, k , of subsamples.

3. Bootstrap approximations

The asymptotic tests based on variance ratios are liable to have important size distortions for several reasons. The distribution of variance ratios is asymmetric because they are bounded by zero from below. The use of large q relative to sample size T can affect the finite sample properties of estimates of Ω_k or Σ_k . The maximum deviation tests based on the Bonferroni inequality are very conservative in some circumstances. The finite sample properties of Box–Pierce and regression-based tests can also be poorly approximated by the asymptotic distribution in situations with a small number of non-overlapping excess returns. Therefore, it is worth pursuing better approximations of the actual joint distribution of those statistics, which also permit a wider range of tests in applications to be conducted.

We use bootstrap techniques to improve the finite sample performance of those serial dependence tests that depend on the joint distribution of subsample statistics and to provide an approximation of the asymptotic distribution of any particular continuous functional. One possibility in the present context is to use the block bootstrap method by Künsch (1989), which allows for the approximation of the asymptotic distribution of statistics based on a weak dependent time series, such as m -dependent series under the nonparametric null hypothesis $\mathcal{H}_0^{(k)}$. However, we instead adopt the parametric bootstrap procedures based on the null $H_0^{(k)}$ under Assumptions 2.1 and 2.2. This approach avoids selecting the order of an approximating parametric model, such as the autoregressive sieve bootstrap of Bühlmann (1997), because the dependence horizon is known in our case.

To conserve space, we only provide a description of our bootstrap procedure for the statistics that are continuous functionals of the subsample variance ratios, $\widehat{VR}_k^{(j)}$, $j = 1, \dots, k$, but similar ideas apply to pooled estimates or other statistics depending on the autocovariances of subsamples.

1. Fit an MA ($k - 1$) model with an intercept to the original sample $S = \{\xi_{1|k}, \xi_{2|k}, \dots, \xi_{T|k}\}$ and obtain residuals \hat{e}_t , $t = 1, 2, \dots, T$ setting the initial values to zero.
2. Obtain an independent resample of size $2T$, $\{\tilde{e}_1^*, \tilde{e}_2^*, \dots, \tilde{e}_{2T}^*\}$ from the empirical distribution of the centered residuals $\tilde{e}_t = \hat{e}_t - \bar{e}_T$, where $\bar{e}_T = T^{-1} \sum_{t=1}^T \hat{e}_t$.
3. Take the moving averages y_t^* of the resampled errors \tilde{e}_t^* from step 2 using the estimated parameter values in step 1 and construct a bootstrap sample $S^* = \{\xi_1^*, \xi_2^*, \dots, \xi_T^*\} = \{y_{T+1}^*, y_{T+2}^*, \dots, y_{2T}^*\}$.
4. Divide the bootstrap sample into k subsamples, $S_1^* = \{\xi_1^*, \xi_{k+1}^*, \dots, \xi_{T-k+1}^*\}, \dots, S_k^* = \{\xi_k^*, \xi_{2k}^*, \dots, \xi_{T-k}^*\}$. Then, calculate variance ratios $\sqrt{T/k}(\widehat{VR}_k^{(j)*}(q) - 1)$, $j = 1, \dots, k$ and construct any test statistic of interest from these.
5. Repeat steps 2 to 4 B times.
6. Obtain estimates of the critical values for the one-sided and two-sided tests based on the empirical distribution of the corresponding bootstrap statistics.

These estimated critical values can be compared to the test statistics obtained from the data. Note that this bootstrap algorithm simulates the distribution of variance ratio statistics under the null $H_0^{(k)}$ by imposing a MA ($k - 1$) structure on the independent resampled residuals. In step 2, we obtain a sample of size $2T$ to eliminate the influence of the initial values, which are set to zero.

The next lemma formally justifies that our bootstrap method can be applied to the statistics introduced in the previous section when the parameter estimates of the invertible MA structure are asymptotically equivalent to the maximum likelihood estimates, following similar procedures to those in Bose (1990) and Kreiss and Franke (1992).

Lemma 3.1. Under $H_0^{(k)}$ and Assumption 2.2, and if the roots of the MA ($k - 1$) polynomial with $c_k = 1$ are outside the unit circle, $c_1 \neq 0$, $\sqrt{T/k}(\widehat{VR}_k^{(j)*}(q) - 1)$, $j = 1, \dots, k$ converges in distribution a.s. to the same asymptotic distribution as $\sqrt{T/k}(\widehat{VR}_k^{(j)}(q) - 1)$, $j = 1, \dots, k$.

Given that $\widehat{V}_{j|k}(1)$ and $\widehat{V}_{j|k}(q)$ have different limits for $q > 1$ under the alternative, the consistency of the bootstrap procedure follows if the roots of the estimated MA ($k - 1$) polynomial are chosen outside the unit circle. Then, the bootstrap distribution converges to a well-defined limit and the estimated quantiles are finite asymptotically.

The same bootstrap method as in Lemma 3.1 can be justified for all tests described in Section 2 based on the sample autocorrelations. These bootstrap procedures can be expected to improve over the asymptotic χ_k^2 distribution of the Wald statistic, $W_k(q)$, and to closely approximate the asymptotic distribution of $Minz_k(q)$ and the related max statistic, replacing the conservative asymptotic critical values based on the Bonferroni inequality. The bootstrap approximation for the finite sample distributions of the minimum and maximum statistics should be able to capture the induced skewness in extreme value distributions. This skewness provides feasible and powerful methods by exploiting right (left) hand side tests based on the maximum (minimum) of individual variance ratios.

To account for the presence of conditional heteroscedasticity in our bootstrap approximation as allowed by Assumption 2.1, we first need to specify a parametric form for the conditional variance of the innovations in the MA specification. For that, we assume that the disturbances e_t follow a flexible GARCH(p_1, p_2) parameterization,

$$e_t = \varepsilon_t \sigma_t \tag{18}$$

with ε_t being iid (0, 1) and

$$\sigma_t^2 = \theta_0 + \sum_{i=1}^{p_1} \theta_{1,i} e_{t-i}^2 + \sum_{j=1}^{p_2} \theta_{2,j} \sigma_{t-j}^2. \tag{19}$$

Note that e_t is a martingale difference sequence, while condition (1) holds for symmetric GARCH processes. Then, we can adapt the first three steps of the previous bootstrap procedure accordingly under appropriate conditions on the stationarity of (19).

1. Fit an MA ($k - 1$)-GARCH(p_1, p_2) model by quasi maximum likelihood with an intercept to the original sample $S = \{\xi_{1|k}, \xi_{2|k}, \dots, \xi_{T|k}\}$ and obtain (standardized) residuals \hat{e}_t , $t = 1, 2, \dots, T$ setting initial values to zero.
2. Obtain an independent resample of size $2T$, $\{\tilde{e}_1^*, \tilde{e}_2^*, \dots, \tilde{e}_{2T}^*\}$, from the empirical distribution of the centered residuals $\tilde{e}_t = \hat{e}_t - \bar{e}_T$, where $\bar{e}_T = T^{-1} \sum_{t=1}^T \hat{e}_t$.
- 3a. Simulate a GARCH(p_1, p_2) series \tilde{e}_t^* of size $2T$ with parameter values from the estimates of step 1 and resampled errors \tilde{e}_t^* from step 2 as innovations.
- 3b. Compute the moving averages y_t^* of the simulated heteroscedastic errors \tilde{e}_t^* from step 3a using the estimated parameter values in step 1 and construct a bootstrap sample $S^* = \{\xi_1^*, \xi_2^*, \dots, \xi_T^*\} = \{y_{T+1}^*, y_{T+2}^*, \dots, y_{2T}^*\}$.

Then, the procedure continues as before in steps 4–6. The justification of the bootstrap methods for the statistics of GARCH processes under regularity conditions on θ and ε_t follows from Hidalgo and Zaffaroni (2007) [see also Assumption A and the discussion in Corradi and Iglesias (2008)], but we omit the details.

4. An application: uncovered interest parity

We now apply the econometric methods developed in the previous sections to the tests of uncovered interest rate parity

(UIP), that is, the expectations hypothesis in foreign exchange markets. Under the assumptions of rational expectations and risk-neutral preferences, UIP is defined by

$$E_t[s_{t+k}] - s_t = i_{t|k} - i_{t|k}^*, \quad \text{for each maturity } k,$$

where $E_t[\cdot]$ denotes the mathematical expectation given the information set available at time t , s_t is the log of the spot exchange rate, or the log of the home currency price of foreign currency at time t , and $i_{t|k}(i_{t|k}^*)$ is the nominal interest rate on home (foreign) deposits with a maturity of k periods. Assuming that covered interest rate parity holds,

$$f_{t|k} - s_t = i_{t|k} - i_{t|k}^*, \quad \text{for each maturity } k,$$

where $f_{t|k}$ is the log of the forward exchange rate, or the time t home currency price of the foreign currency delivered at time $t+k$. Then, UIP is equivalent to the unbiasedness hypothesis of forward exchange rates defined by

$$E_t[s_{t+k}] = f_{t|k}, \quad \text{for each maturity } k. \tag{20}$$

In this sense, UIP implies that the foreign excess return between t and $t+k$, $s_{t+k} - f_{t|k}$, should be unpredictable using any variables in the time t information set. As the definition of the foreign excess return indicates, any predictability tests need to take into account the $(k-1)$ -dependence of excess returns, which might motivate the use of our sample splitting methods.

The alternative hypothesis we are interested in has the following form:

$$E_t[s_{t+k}] = f_{t|k} - p_{t|k}, \quad \text{for each maturity } k, \tag{21}$$

where $p_{t|k}$ is a deviation from UIP interpreted as a risk premium or as an expectational error. In this paper, we mainly consider these two alternatives because they have been widely used in the literature.¹² Furthermore, Moon and Velasco (2011) argue that these two alternatives in the literature tend to generate the opposite sign of serial dependence of excess returns, which can be used to judge the performance of economic models. For example, they show that the rational expectations risk premium models generate negative serial dependence patterns, while the models of expectational errors tend to generate positive serial dependence patterns.

As a data-generating process, we use the typical monetary model of exchange rates.¹³ In the model, the home money market relationship is given by

$$\ln M_t = \ln P_t + \gamma_y \ln Y_t - \phi_k i_{t|k}, \tag{22}$$

where M , P , and Y are the home money supply, the price level, and the output, respectively. γ_y is the income elasticity of money demand and ϕ_k is the interest semi-elasticity of money demand, which varies with maturity k . We assume that a similar equation holds in the foreign country. The corresponding foreign variables are denoted by asterisks and the parameters of the money demand are the same in both countries. From (21), covered interest parity, the home money market relationship (22) and its foreign counterpart, we derive a setup for the determination of the exchange rate:

$$s_t = bE_t[s_{t+k}] + bp_{t|k} + (1-b)w_t + (1-b)\varpi_t, \tag{23}$$

where $b = \frac{\phi_k}{1+\phi_k}$ is the discount factor, w_t is the linear combination of the fundamental variables, $w_t = \ln M_t - \ln M_t^* - \gamma_y(\ln Y_t - \ln Y_t^*)$, and ϖ_t is the log of the real exchange rate defined by $\varpi_t = s_t + \ln P_t^* - \ln P_t$. Eq. (23) implies that the model generates spot rates of $s_t, s_{t+k}, s_{t+2k}, \dots$. Therefore, when $k > 1$, we can obtain k populations that correspond to k subsamples in Section 2. Assuming PPP holds, the “no-bubbles” solution to Eq. (23) is

$$s_t = (1-b) \sum_{i=0}^{\infty} b^i E_t[w_{t+ik}] + b \sum_{i=0}^{\infty} b^i E_t[p_{t+ik|k}]. \tag{24}$$

In this present value model, the spot exchange rate s_t is expressed as the discounted sum of the current and expected future fundamentals as well as deviations from UIP.

For the Monte Carlo simulations in the next section, we use four different models. One model assumes UIP and the other three specifications consider a deviation from UIP: one for the rational expectations risk premium and the others for the expectational errors. Although all of the models considered in this section share the setup for the exchange rate (23) and the present value relationship (24), they are different in terms of modeling the deviation from UIP.

To generate s_t , we need to model the processes for w_t and $p_{t|k}$, which is beyond the scope of the current paper. Instead, we present reduced form expressions for those processes. We begin with the process for fundamentals, w_t , which is assumed to be identical in all four models. We choose a random walk model for w_t ¹⁴

$$w_t = w_{t-1} + e_t, \tag{25}$$

where e_t is $iid(0, \sigma^2)$. Here, we assume that the process for w_t is formulated at, for example, weekly frequency; while adhering to the UIP conditions, all of the models for the spot and forward exchange rates are built at “ k -week” frequency.

Assume UIP holds ($p_{t|k} = 0$). Then, from Eqs. (24) and (25), we obtain

$$s_t = w_t.$$

Because Eqs. (20) and (25) imply $f_{t|k} = w_t$, the foreign excess return between t and $t+k$ is

$$s_{t+k} - f_{t|k} = \sum_{j=1}^k e_{t+j}. \tag{26}$$

We call this result Model 1 in our Monte Carlo simulations in the next section.

We present a risk premium alternative in Model 2. The process for the time varying risk premium between t and $t+k$ is given by

$$p_{t|k} = (1 - \varphi_k)\bar{p} + \varphi_k p_{t-k|k} + v_t + \dots + v_{t-k+1}, \tag{27}$$

where $0 < \varphi_k < 1$ and v_t is $iid(0, \sigma_v^2)$. The process for the risk premium is modeled such that it conforms with a maturity k in Eq. (21). Using Eqs. (24), (25) and (27), the expression for the spot exchange rate is obtained as

$$s_t = w_t + \frac{b}{1 - b\varphi_k} p_{t|k}, \tag{28}$$

¹² There are other explanations in the literature that are mainly related to small sample problems such as the peso problem, learning, and statistical biases. The variance ratio tests employed in the paper are robust to the statistical biases that typically arise in the regression-based tests.

¹³ See, e.g., Engel and West (2005) and Obstfeld and Rogoff (2002) for the rational expectations models and Frankel and Froot (1990) for the expectational errors models. See also the references therein.

¹⁴ Several processes for fundamentals, w_t , have been used in the literature, although the particular use depends on the objective of the study. For example, Tauchen (2001) uses a stationary AR(1) model, Engel and West (2005) consider an integrated AR(1) model, and Baillie and Bollerslev (2000) assume a fads model which is the sum of random walk and stationary AR(1) components. Obviously, the size of the serial dependence tests used would be identical in those fundamental processes. Furthermore, the relative power performance among the aggregation methods introduced in the previous sections remains unchanged in these fundamental processes. So, we choose the random walk model for simplicity.

where the constant terms are omitted for simplicity. Here, the spot exchange rate is expressed by the sum of a random walk fundamental and a stationary risk premium, mirroring a well-known fads model for studying the long-run predictability of stock returns in Fama and French (1988) and Poterba and Summers (1988). The forward exchange rate is derived from Eqs. (21), (25), (27) and (28)

$$f_{t|k} = E_t[s_{t+k}] + p_{t|k} = w_t + \frac{1}{1 - b\varphi_k} p_{t|k}.$$

Then, the foreign excess return between t and $t + k$ under the risk premium alternative is

$$s_{t+k} - f_{t|k} = \sum_{i=1}^k e_{t+i} + \frac{b}{1 - b\varphi_k} \sum_{i=1}^k v_{t+i} - p_{t|k}, \tag{29}$$

where the first two terms in the right-hand side of Eq. (29) are rational forecasting errors: e_t is from the fundamental process and v_t is from the risk premium process. Eq. (29) can be viewed as a reduced form expression for the excess return that is derived from the time-varying risk premium models in the literature. The expression shows that the forecasting errors will be correlated with the future values of the risk premium, reflecting a feedback mechanism that mainly determines the sign of the autocorrelations of excess returns. As shown in Moon and Velasco (2011), this model tends to generate a negative autocorrelation of excess returns for a reasonable range of parameter values.¹⁵

We now present an expectational error alternative based on Frankel and Froot (1990) that generates a positive autocorrelation of excess returns. There are three types of agents. One type is portfolio managers who participate in currency transactions. The other two, fundamentalists and noise traders (chartists), merely issue the forecasts of future exchange rates and do not participate in the transactions. The portfolio managers' expectation, which equals the market expectation, is given by a weighted linear combination of the forecasts of the other two agents

$$E_t^m[s_{t+k}] = (1 - \lambda)E_t[s_{t+k}] + \lambda E_t^n[s_{t+k}], \tag{30}$$

where $E_t[\cdot]$ is the expectation of fundamentalists whose expectation is rational, $E_t^n[\cdot]$ is the expectation of noise traders, and $0 \leq \lambda \leq 1$. We assume that the noise traders' expectations are regressive toward a long-run equilibrium exchange rate, \bar{s}_t ,

$$E_t^n[s_{t+k}] = (1 - g)s_t + g\bar{s}_t, \tag{31}$$

where \bar{s}_t is the difference in the consumer price indexes between domestic and foreign countries and $0 \leq g < 1$ is the adjustment speed of s_t towards \bar{s}_t .¹⁶ We denote the k -week real exchange rate $\varpi_{t|k} = s_t - \bar{s}_t$, which is assumed to follow a stationary process,

$$\varpi_{t|k} = (1 - \psi_k)\bar{\varpi} + \psi_k\varpi_{t-k|k} + \eta_t + \dots + \eta_{t-k+1}, \tag{32}$$

where $\bar{\varpi}$ is the constant long-run level of the real exchange rate, $0 < \varphi_k < 1$, and η_t is $iid(0, \sigma_\eta^2)$. Analogous to the risk premium process, the process for the real exchange rate is modeled such

¹⁵ We restrict to $0 < \varphi_k < 1$ following the convention in the literature. However, as shown in Moon and Velasco (2011), the results would go through even when $\varphi_k = 0$. Furthermore, $\varphi_k = 1$ will provide additional information for the identification because it implies that the autocovariance of excess returns between time t and $t + q$ is not reverting toward zero as q increases and is always the same as that between t and $t + 1$. Furthermore, when $\varphi_k < 0$, the sign of the autocorrelations oscillates, which can also be used for the identification.

¹⁶ Frankel and Froot (1987) presented several empirically-relevant formulations for the noise traders' expectations such as distributed lag expectations and adaptive expectations. We choose one of them.

that it conforms with a maturity k . We also assume that there is no market risk premium so that

$$f_{t|k} = E_t^m[s_{t+k}] = E_t[s_{t+k}] + p_{t|k}^e \quad \text{for each } k, \tag{33}$$

where $p_{t|k}^e = E_t^m[s_{t+k}] - E_t[s_{t+k}]$ is an expectational error due to the presence of noise traders and represents a deviation from UIP. Then, analogous to the risk premium model, we have a setup for determining the exchange rate under the expectational error alternative using

$$s_t = bE_t[s_{t+k}] + bp_{t|k}^e + (1 - b)w_t + (1 - b)\varpi_{t|k}.$$

Note that the main difference from the previous setup under the rational expectations risk premium is that the risk premium $p_{t|k}$ is replaced by the expectational error. Using the definition of $p_{t|k}^e$ and Eqs. (30)–(32), we can rewrite the above equation as

$$s_t = \bar{b}E_t[s_{t+k}] + \bar{b}p_{t|k}^e + (1 - \bar{b})w_t, \tag{34}$$

where $\bar{b} = \frac{b(1-\lambda)}{1-b\lambda}$ and $\bar{p}_{t|k}^e = \frac{1-b(1+\lambda g)}{b(1-\lambda)}\varpi_{t|k}$. The discount factor, \bar{b} , is now related not only to the interest semi-elasticity of the money demand but also to the weight of the noise traders' expectation, λ , in the market expectation.

Assuming no-bubble solutions, the foreign excess return is derived from Eqs. (25) and (30)–(34)

$$s_{t+k} - f_{t|k} = \sum_{i=1}^k e_{t+i} + \frac{1 - b(1 + \lambda g)}{1 - b\lambda - b(1 - \lambda)\psi_k} \sum_{i=1}^k \eta_{t+i} - p_{t|k}^e. \tag{35}$$

For the sake of simplicity, we relegate the derivation of Eq. (35) to Appendix B. We call this Model 3 in the simulations. Analogous to the risk premium alternative, the forecasting errors are correlated with the future values of $p_{t|k}^e$, illustrating the feedback mechanism. However, the excess returns now exhibit positive autocorrelations for a reasonable range of parameter values.¹⁷

The power pattern of the variance ratio test with q would be quite similar between the risk premium and the expectational errors alternatives because both alternatives follow stationary AR(1) processes, although rejections mainly occur at the opposite tail. That is, the power of the test will initially increase and then decrease with q [see, e.g., Lo and MacKinlay (1989)]. For comparison, we consider another alternative that generates a different power pattern with q by modifying the noise traders' expectation in Eq. (31) in the following way: $E_t^n[s_{t+k}] = E_{t-k}[s_t]$. Assuming PPP holds, then the foreign excess return between t and $t + k$ is

$$s_{t+k} - f_{t|k} = (1 - b\lambda) \sum_{i=1}^k e_{t+i} + \lambda \sum_{i=0}^{k-1} e_{t-i}. \tag{36}$$

Again, we relegate the derivation to Appendix B. We call this specification Model 4. As shown in the Monte Carlo experiments, Eq. (36) generates not only positive autocorrelations but also a uniformly declining power of the serial dependence tests with q .

5. Monte Carlo simulations

We conduct Monte Carlo experiments to study the finite sample properties of the test statistics for m -dependent data developed in Section 2. We explore the properties of both asymptotic and parametric bootstrap tests for each statistic. To improve the numerical efficiency in the simulations of bootstrap asymptotic size and power, we use the method of Giacomini et al. (2012), where each simulation generates only one bootstrap resample and a single critical value is estimated from all of the resamples.

¹⁷ Here, we confine our attention to the case where the real exchange rate is mean reverting in the long run. However, allowing $\psi_k = 1$ would only strengthen the result.

5.1. Econometric frameworks for Monte Carlo simulations

To measure the size and power of the test statistics, we use four models presented in Section 4:

- Model 1 uses Eqs. (25) and (26).
- Model 2 uses Eqs. (25), (27) and (29).
- Model 3 uses Eqs. (25), (32) and (35).
- Model 4 uses Eqs. (25) and (36).

Model 1 generates excess returns under UIP, so the rejection rates provide the empirical size of the test statistics. The remaining models generate excess returns, which exhibit either negative or positive serial dependence. The rejection rates from these models measure the power of the tests. Model 2 generates a negative serial dependence of excess returns, while Model 3 and Model 4 generate a positive serial dependence but with different power patterns over the aggregation value q .

Our simulations use the following parameterization. For each of the four models, we consider two specifications for e_t in Eq. (25). One specifies that e_t follows an *iid* normal distribution with mean zero. The other specification assumes that e_t follows the process in Eqs. (18)–(19) with $p_1 = 1$, $p_2 = 0$ and $\theta_{1,1} = 0.5$, which allows for the conditional dependence in the fourth moment of e_t . The sample size for each simulation is $T = 33 * 52$, which corresponds to the currently available sample size in weekly floating exchange rates. k is set at 13 so that it represents one quarter. The quarterly interest semi-elasticity of money demand ϕ_k is set at 20. We set φ_k in Eq. (27) at 0.81, based on the median estimate of the first order autocorrelations of the three-month forward premium in our sample in the next section. We set ψ_k in Eq. (32) at 0.95, using the median value of the first order autocorrelations of the U.S. bilateral quarterly real exchange rates in our sample. The weight λ is set at 0.3 and the speed of adjustment g is set at 0.25, following the estimation results of Frankel and Froot (1987). The correlation between e_t and v_t is set at 0 in Model 2 and the correlation between e_t and η_t is set at 0.5 in Model 3. We assume $\sigma = \sigma_v = \sigma_\eta$ and set so that the variance of the excess return broadly matches the data in the next section. With these parameter values, the present value model generates spot and forward exchange rates whose time series properties, in terms of persistency and volatility, are broadly consistent with the data.

5.2. Simulation results

Tables 1 and 2 report the results of the simulation experiments from Model 1, while Tables 3–5 report the results from Models 2, 3, and 4, respectively. Table 1 reports the test results based on the asymptotic critical values, while the other tables report the results based on the critical values from the parametric bootstrap empirical distribution constructed using the procedures in Section 3. Panel A of each table reports the results from the models in which e_t is *iid*, while Panel B reports those from the models in which e_t is conditionally heteroscedastic. We conduct statistical tests at conventional significance levels against both the right-tail and left-tail alternatives but only report the results of the tests at the 5% significance level to conserve space. The results in the tables are the rejection rates obtained from 10,000 simulations. The range of aggregation values is set such that the maximum value of q is 10 years relative to a base period of a quarter and includes 2, 4, 8, 12, 16, 20, 32, and 40 quarters. For comparison, we also set $n = q$, the holding period horizon in the Fama–French regression, except that $q = 2$ is replaced with $n = 1$.

5.2.1. Size

Panel A in Table 1 reports the rejection rates of the serial dependence tests based on asymptotic critical values for the

iid excess returns. We use three types of serial dependence tests: variance ratio, Box–Pierce portmanteau, and Fama–French regression tests. For the variance ratio tests, we use several aggregates based on the pooled method, Bonferroni bounds, and the Wald method. Overall, most tests have a reasonable size at the right-tail for the smaller aggregation value q , while the variance ratio tests under-reject at the left-tail.

We begin with the test results at the right-tail. The empirical sizes of the t -statistic of pooled variance ratios, $z_k(q)$, appear to be reasonable at the right tail over all q considered, although the test slightly over-rejects for large q . For example, the rejection rates associated with the aggregation values $q = 2, 4, 8, 12, 16, 20, 32$, and 40 quarters, are 4, 5, 6, 7, 7, 7, 8, and 9% at the right-tail, respectively. The size of the Bonferroni maximum variance ratio test is close to the nominal value even for large q : its rejection rates are approximately 5% for $q = 32$ and 40. The right-tail t -test from the Fama–French regression also has a reasonable size over all q , while the Box–Pierce pooled statistic tends to slightly over-reject for large q .

However, the empirical sizes of the t -statistics of the pooled variance ratios become distorted for large q at the left-tail. For example, the rejection rates of the left-tail test are 2% and 1% for $q = 32$ and $q = 40$ quarters, respectively. Similarly, the Bonferroni minimum variance ratio test appears not to reject at all over most values of q . As in Richardson and Stock (1989), one possible reason for this size distortion can be that the variance ratios become inconsistent for large q relative to the sample size T .

Panel B in Table 1 reports the rejection rates of the serial dependence tests in the presence of conditional heteroscedasticity in excess returns. All of the tests produce quite similar rejection patterns to those in Panel A.

Table 2 reports the rejection rates of the serial dependence tests based on the critical values obtained from the parametric bootstrap method. In contrast to the asymptotic tests, there are no size distortions at both tails even for large q . The empirical sizes of all the tests are close to their nominal value at both tails and for all q considered. For example, the rejection rates of the t statistics of estimated pooled variance ratios are all 5% at both tails for each aggregation value q . The results are almost identical even in the presence of conditional heteroscedasticity.

From now on, we will focus on the results from the parametric bootstrap method because it corrects the potential size distortions of the serial dependence tests from both skewness and Bonferroni inequality. Furthermore, we will mainly discuss the results from the models with *iid* e_t because the tests produce similar size and power properties for both specifications of e_t .¹⁸

5.2.2. Power

We now discuss the power properties of the serial dependence tests. In general, the power of the tests is sensitive to the parameterization of the simulated models. However, the two important features that we are interested in, the sign of the autocorrelations of the excess returns and the power pattern over q , are not sensitive to a broad range of parameter values. Nor is the relative performance over tests considered. Therefore, we do not provide further sensitivity analysis on the power of the tests with respect to changes in the parameter values.

¹⁸ We also conduct the rank-based variance ratio tests because their size and power properties are not known for the m -dependent time series data, although their size is exact in finite samples for the 0-dependence data as in Wright (2000). We find that all three aggregates of rank-based variance ratios such as the minimum, maximum, and median have the size close to the nominal value. These results are available upon request.

Table 1
The size of the tests based on asymptotic critical values: Model 1.

$k = 13$									
q	Variance ratio					Box–Pierce		Fama–French	
	Pooled		Bonferroni max	Bonferroni min	Wald	Pooled	Regression		
	z_k		$Maxz_k$	$Minz_k$	W_k	\hat{Q}_k	$t_{\hat{\beta}_{n,k}}$		
5%-L		5%-R	5%-R	5%-L	5%-R	5%-R	5%-L	5%-R	
Panel A. e_t is iid Normal									
2	0.08	0.04	0.02	0.02	0.10	0.06	0.11	0.06	
4	0.08	0.05	0.03	0.00	0.07	0.08	0.17	0.05	
8	0.07	0.06	0.03	0.00	0.04	0.10	0.24	0.05	
12	0.05	0.07	0.04	0.00	0.03	0.11	0.30	0.04	
16	0.05	0.07	0.04	0.00	0.02	0.11	0.38	0.04	
20	0.04	0.07	0.04	0.00	0.02	0.12	0.44	0.04	
32	0.02	0.08	0.05	0.00	0.02	0.12	0.62	0.06	
40	0.01	0.09	0.05	0.00	0.02	0.11	0.69	0.08	
Panel B. e_t is conditional heteroscedastic									
2	0.08	0.04	0.03	0.02	0.14	0.06	0.11	0.06	
4	0.07	0.05	0.04	0.00	0.10	0.06	0.16	0.05	
8	0.06	0.06	0.04	0.00	0.07	0.06	0.22	0.05	
12	0.05	0.07	0.04	0.00	0.06	0.06	0.28	0.04	
16	0.04	0.07	0.04	0.00	0.05	0.06	0.36	0.04	
20	0.03	0.08	0.05	0.00	0.05	0.06	0.43	0.04	
32	0.01	0.08	0.05	0.00	0.04	0.05	0.62	0.06	
40	0.00	0.09	0.05	0.00	0.03	0.04	0.69	0.08	
T	1716	1716	1716	1716	1716	1716	1716	1716	

Notes: The notes presented here are applied to Tables 1–5. Each table displays rejection rates of the serial dependence tests at the 5% significant level, obtained from 10,000 simulations. The rejection rates range between 0 and 1 including bounds. Table 1 reports the results using the conventional asymptotic critical values, while Tables 2–5 report the results using bootstrap critical values based on the parametric bootstrap method. Panel A reports the results from the models in which e_t follows an iid Normal process, while Panel B reports those in which e_t follows Eqs. (18)–(19) with $p_1 = 1$, $p_2 = 0$ and $\theta_{1,1} = 0.5$. ‘5%-L’ represents the rejection rates of the 5% left-tail test and ‘5%-R’ represents those of the 5% right-tail test. $k = 13$ represents a maturity of 13 weeks (or one quarter). q is an aggregation value which represents a holding period of quarters relative to a base period of one-quarter. T represents sample size of weekly observations.

The t statistic of pooled variance ratios, z_k , is defined in (10); the t statistics of Bonferroni maximum/minimum variance ratios, $Maxz_k$ and $Minz_k$, are defined in (8); the Wald variance ratio statistic, W_k , is defined in (7) with $R = I_k$; the Box–Pierce pooled statistic, \hat{Q}_k , is defined in (16); and the t statistic of $\hat{\beta}_{n,k}$ from the Fama–French regression (11) is obtained calculating the Newey and West standard errors with $n * k - 1$ lag lengths. We set $n = q$ except that $q = 2$ is replaced with $n = 1$. In Tables 2–5, the t statistics of the maximum/minimum variance ratios, $Maxz_k$ and $Minz_k$, are defined in (8); the Box–Pierce maximum statistic, $MaxQ_k$, is defined in (15). We slightly abuse the notation for the t statistics of the maximum/minimum variance ratios, $Maxz_k$ and $Minz_k$, to save space. We also define the t statistic of the median variance ratios by $Medz_k(q) = median\{U_k(q)\}$ based on the ordered statistics of $U_k(q)$ in (4).

Table 2
The size of the tests based on the parametric bootstrap method: Model 1.

$k = 13$													
q	Variance ratio								Box–Pierce		Fama–French		
	Pooled		Median	Max	Min	Wald	Pooled	Max	Regression				
	z_k		$Medz_k$	$Maxz_k$	$Minz_k$	W_k	\hat{Q}_k	$MaxQ_k$	$t_{\hat{\beta}_{n,k}}$				
5%-L		5%-R	5%-L	5%-R	5%-L	5%-R	5%-L	5%-R	5%-R	5%-L	5%-R		
Panel A. e_t is iid Normal													
2	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
4	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
8	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
12	0.06	0.05	0.06	0.05	0.06	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05
16	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.05
20	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.06
32	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.05	0.06
40	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.05	0.04	0.05	0.05	0.05	0.05
Panel B. e_t is conditional heteroscedastic													
2	0.05	0.05	0.06	0.05	0.06	0.04	0.05	0.05	0.05	0.06	0.05	0.05	0.05
4	0.05	0.05	0.06	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05
8	0.06	0.05	0.06	0.05	0.06	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05
12	0.06	0.05	0.06	0.05	0.06	0.04	0.05	0.06	0.04	0.05	0.05	0.05	0.05
16	0.06	0.05	0.05	0.05	0.06	0.04	0.05	0.05	0.04	0.05	0.05	0.05	0.05
20	0.05	0.05	0.05	0.05	0.06	0.04	0.04	0.05	0.04	0.05	0.05	0.05	0.04
32	0.05	0.05	0.05	0.05	0.06	0.04	0.04	0.05	0.04	0.05	0.05	0.05	0.04
40	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.05	0.04	0.05	0.05	0.05	0.05
T	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716

Negative serial dependence

Table 3 reports the results of the serial dependence tests from Model 2. The rejections mainly occur at the left-tail and the

simulated variance ratios are less than 1, suggesting that the excess returns generated from Model 2 exhibit negative autocorrelations. Furthermore, the tests produce a hump-shaped power pattern over

Table 3

The power of the tests based on the parametric bootstrap method: Model 2.

$k = 13$													
q	Variance ratio								Box–Pierce		Fama–French		
	Pooled		Median		Max		Min		Wald	Pooled	Max	Regression	
	z_k		$Medz_k$		$Maxz_k$		$Minz_k$		W_k	\hat{Q}_k	$MaxQ_k$	$t_{\hat{\beta}_{n,k}}$	
	5%-L	5%-R	5%-L	5%-R	5%-L	5%-R	5%-L	5%-R	5%-R	5%-R	5%-R	5%-L	5%-R
Panel A. e_t is iid Normal													
2	0.29	0.00	0.29	0.00	0.27	0.00	0.26	0.00	0.03	0.28	0.19	0.29	0.00
4	0.45	0.00	0.44	0.00	0.43	0.00	0.43	0.00	0.02	0.27	0.19	0.45	0.00
8	0.60	0.00	0.60	0.00	0.59	0.00	0.58	0.00	0.01	0.23	0.16	0.40	0.00
12	0.67	0.00	0.67	0.00	0.66	0.00	0.66	0.00	0.01	0.20	0.15	0.28	0.00
16	0.70	0.00	0.71	0.00	0.71	0.00	0.69	0.00	0.00	0.19	0.13	0.19	0.00
20	0.71	0.00	0.70	0.00	0.71	0.00	0.70	0.00	0.00	0.17	0.13	0.14	0.00
32	0.64	0.00	0.64	0.00	0.65	0.00	0.62	0.00	0.00	0.17	0.12	0.06	0.01
40	0.58	0.00	0.58	0.00	0.59	0.00	0.57	0.00	0.00	0.16	0.12	0.03	0.01
Panel B. e_t is conditional heteroscedastic													
2	0.29	0.00	0.26	0.00	0.26	0.00	0.23	0.00	0.02	0.25	0.16	0.28	0.00
4	0.42	0.00	0.40	0.00	0.40	0.00	0.38	0.00	0.01	0.23	0.15	0.44	0.00
8	0.59	0.00	0.56	0.00	0.56	0.00	0.55	0.00	0.01	0.19	0.13	0.40	0.00
12	0.67	0.00	0.65	0.00	0.65	0.00	0.62	0.00	0.00	0.17	0.12	0.30	0.00
16	0.69	0.00	0.66	0.00	0.67	0.00	0.64	0.00	0.00	0.16	0.11	0.20	0.00
20	0.69	0.00	0.66	0.00	0.66	0.00	0.63	0.00	0.00	0.16	0.11	0.15	0.00
32	0.63	0.00	0.61	0.00	0.61	0.00	0.57	0.00	0.00	0.14	0.11	0.05	0.01
40	0.58	0.00	0.56	0.00	0.56	0.00	0.53	0.00	0.00	0.14	0.11	0.03	0.01
T	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716

q : the rejection rates initially increase and then decrease with q . For example, the rejection rates of a pooled variance ratio test associated with the aggregation values $q = 2, 4, 8, 12, 16, 20, 32$, and 40 quarters, are 30, 44, 58, 67, 70, 70, 65, and 60% at the left-tail, respectively. The presence of the strongly persistent component (the risk premium) in the exchange rates generated from Model 2 explains this non-monotonic power pattern of the variance ratio tests consistent with Lo and MacKinlay (1989), who show that a mean reverting component in asset prices generates this nonmonotone power pattern.

We compare the power of three serial dependence tests. The variance ratio tests, except for the Wald method, are more powerful than the Box–Pierce and Fama–French regression tests in that the power of the former is greater than those of the latter for each q . Furthermore, the maximum power of the former is much greater than the latter. For example, the largest rejection rate of the variance ratio tests over q is approximately 70%, while those of the Box–Pierce portmanteau and Fama–French t -tests are approximately 29% and 45%, respectively. The variance ratio tests based on the pooled, median, maximum, and minimum methods have similar power properties in terms of rejection rates and power patterns, while the Wald method performs much worse, with much less power for each aggregation value q . For example, the rejection rates of the Wald variance test associated with the aggregation values are all under 3%.

Positive serial dependence

Table 4 reports the results from Model 3. Here, the rejections mainly occur at the right-tail and the simulated variance ratios are greater than one, suggesting that the excess returns generated from Model 3 display positive autocorrelations. Furthermore, the power of the tests initially increases and then decreases with q . This non-monotonic power pattern is produced because the expectational error is persistent in Model 3, like the risk premium in Model 2.

Similar to Model 2, the variance ratio tests, except for the Wald method, are more powerful than the Box–Pierce and Fama–French regression tests. For example, the maximum power of the variance ratio tests is approximately 47%, while those of the Box–Pierce

portmanteau and Fama–French t -tests are approximately 18% and 36%, respectively. Again, the variance ratio tests based on the pooled, median, maximum, and minimum methods have similar power properties.

Table 5 reports the results from Model 4 which also generates the positive autocorrelations of excess returns but a different power pattern from Model 3. All of the serial dependence tests reject the null most strongly when $q = 2$ and then their power uniformly decreases with q . For example, the rejection rates of the pooled variance ratio test associated with the aggregation values are 100, 87, 50, 34, 26, 22, 14, and 13% at the right-tail. These results suggest that not only the sign of serial dependence but also the power pattern over q can be used to identify a particular economic alternative.

In contrast to Model 2 and Model 3, the Box–Pierce tests are now more powerful than the variance ratio and Fama–French regression tests. Furthermore, the power of the former is decreasing relatively more slowly over q . For example, the rejection rates of the Box–Pierce maximum statistic, $MaxQ_k$, associated with the aggregation values are 99, 99, 92, 84, 77, 71, 60, and 55% at the right-tail, while the rejection rate of the Fama–French test is already approximately 10% for $q = 4$.

6. Tests for the expectations hypothesis in foreign exchange markets

In this section, we use our serial dependence tests for testing the predictability of three-month foreign excess returns from January 1975 to December 2007 using 1716 weekly observations.¹⁹ The log foreign excess returns $s_{t+k} - f_{t|k}$ are measured over a holding period of $k = 13$ weeks and annualized by $(s_{t+k} - f_{t|k}) * 5200/k$. Our sample includes weekly spot prices of the U.S. dollar against the German mark, the British pound, and the Japanese yen as well as three-month prices (forward exchange rates) of the U.S. dollar.

¹⁹ Wednesday's closing price is selected to form our sample. If the following Wednesday is missing, then Thursday's price is used (or Tuesday's if Thursday's is missing).

Table 4
The power of the tests based on the parametric bootstrap method: Model 3.

$k = 13$													
q	Variance ratio									Box–Pierce		Fama–French	
	Pooled		Median		Max		Min		Wald	Pooled	Max	Regression	
	z_k		$Medz_k$		$Maxz_k$		$Minz_k$		W_k	\hat{Q}_k	$MaxQ_k$	$t_{\hat{\rho}_{m,k}}$	
	5%-L	5%-R	5%-L	5%-R	5%-L	5%-R	5%-L	5%-R	5%-R	5%-R	5%-R	5%-L	5%-R
Panel A. e_t is iid Normal													
2	0.01	0.28	0.01	0.28	0.01	0.25	0.01	0.25	0.09	0.10	0.15	0.01	0.26
4	0.01	0.37	0.01	0.37	0.01	0.34	0.01	0.36	0.09	0.10	0.16	0.01	0.36
8	0.01	0.45	0.01	0.45	0.01	0.42	0.01	0.45	0.12	0.10	0.18	0.01	0.33
12	0.01	0.47	0.01	0.47	0.01	0.43	0.01	0.47	0.16	0.09	0.17	0.01	0.26
16	0.01	0.47	0.01	0.47	0.01	0.43	0.01	0.47	0.21	0.08	0.16	0.02	0.21
20	0.01	0.46	0.01	0.47	0.01	0.43	0.01	0.47	0.22	0.07	0.15	0.02	0.16
32	0.01	0.43	0.01	0.43	0.01	0.40	0.01	0.44	0.23	0.05	0.13	0.04	0.11
40	0.01	0.41	0.01	0.41	0.01	0.38	0.01	0.42	0.23	0.04	0.12	0.07	0.10
Panel B. e_t is conditional heteroscedastic													
2	0.01	0.26	0.01	0.26	0.01	0.22	0.01	0.25	0.11	0.10	0.15	0.01	0.24
4	0.01	0.34	0.01	0.34	0.01	0.30	0.01	0.33	0.10	0.14	0.16	0.01	0.34
8	0.01	0.41	0.01	0.42	0.01	0.37	0.01	0.42	0.11	0.15	0.16	0.01	0.32
12	0.01	0.44	0.01	0.44	0.01	0.39	0.01	0.44	0.12	0.14	0.16	0.02	0.25
16	0.01	0.44	0.01	0.44	0.01	0.40	0.01	0.44	0.14	0.13	0.15	0.02	0.19
20	0.01	0.44	0.01	0.43	0.01	0.39	0.01	0.44	0.13	0.12	0.14	0.03	0.14
32	0.01	0.41	0.01	0.41	0.01	0.37	0.01	0.42	0.15	0.10	0.12	0.04	0.11
40	0.01	0.40	0.01	0.40	0.01	0.35	0.01	0.40	0.16	0.09	0.11	0.07	0.10
T	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716

Table 5
The power of the tests based on the parametric bootstrap method: Model 4.

$k = 13$													
q	Variance ratio									Box–Pierce		Fama–French	
	Pooled		Median		Max		Min		Wald	Pooled	Max	Regression	
	z_k		$Medz_k$		$Maxz_k$		$Minz_k$		W_k	\hat{Q}_k	$MaxQ_k$	$t_{\hat{\rho}_{m,k}}$	
	5%-L	5%-R	5%-L	5%-R	5%-L	5%-R	5%-L	5%-R	5%-R	5%-R	5%-R	5%-L	5%-R
Panel A. e_t is iid Normal													
2	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.14	1.00	0.99	0.00	1.00
4	0.00	0.91	0.00	0.91	0.00	0.87	0.00	0.92	0.04	0.99	0.99	0.02	0.10
8	0.00	0.63	0.00	0.63	0.00	0.57	0.00	0.66	0.01	0.85	0.92	0.04	0.06
12	0.00	0.49	0.00	0.48	0.00	0.43	0.00	0.51	0.02	0.61	0.84	0.04	0.06
16	0.00	0.39	0.00	0.39	0.00	0.34	0.00	0.42	0.02	0.40	0.77	0.04	0.06
20	0.00	0.34	0.00	0.33	0.01	0.29	0.00	0.36	0.03	0.26	0.71	0.04	0.06
32	0.01	0.24	0.01	0.24	0.01	0.20	0.00	0.26	0.03	0.10	0.60	0.06	0.06
40	0.01	0.21	0.01	0.21	0.01	0.18	0.00	0.23	0.04	0.07	0.55	0.05	0.05
Panel B. e_t is conditional heteroscedastic													
2	0.00	1.00	0.00	1.00	0.00	0.99	0.00	1.00	0.13	0.99	0.98	0.00	1.00
4	0.00	0.91	0.00	0.90	0.00	0.84	0.00	0.91	0.03	0.98	0.95	0.01	0.15
8	0.00	0.65	0.00	0.62	0.00	0.54	0.00	0.66	0.01	0.89	0.86	0.04	0.07
12	0.00	0.48	0.00	0.46	0.00	0.39	0.00	0.50	0.01	0.76	0.79	0.04	0.05
16	0.00	0.39	0.00	0.38	0.00	0.31	0.00	0.42	0.01	0.62	0.72	0.04	0.05
20	0.00	0.33	0.00	0.32	0.01	0.27	0.00	0.35	0.01	0.48	0.68	0.05	0.05
32	0.01	0.24	0.00	0.24	0.01	0.20	0.00	0.26	0.01	0.26	0.57	0.05	0.05
40	0.01	0.22	0.01	0.21	0.01	0.18	0.00	0.23	0.02	0.19	0.53	0.05	0.05
T	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716	1716

The data are simultaneously collected from London close bid and ask prices and obtained from Global Insight. The mid prices are used for the empirical study. These are major currencies in foreign exchange markets and have been widely used for testing UIP, the expectations hypothesis in foreign exchange markets.

Table 6 reports the p -values of the test statistics obtained from the parametric empirical bootstrap distributions. The table also reports the pooled variance ratios, $\hat{V}R_k(q)$, defined in Eq. (9).²⁰ Overall, the expectations hypothesis in the foreign

exchange markets is rejected for the three-month excess returns, confirming the previous empirical evidence on the predictability of foreign excess returns. In addition, the patterns of variance ratios over the aggregation values suggest that the foreign excess returns exhibit positive autocorrelations.

The variance ratio tests reject the expectations hypothesis at the 5% level against the right-tail alternative up to the aggregation values $q = 12$ to 16 quarters relative to a three-month base period. Furthermore, the p -values initially decrease and then increase or they tend to be almost zero up to a certain q and then start to increase. For example, the right-tail p -values of the minimum variance ratio test associated with aggregation values $q = 2, 4, 8,$

²⁰ The variance ratios are slightly different among the aggregation methods. For brevity, we only report the pooled variance ratios.

Table 6
Predictability of three-month foreign excess returns: *p*-values.

<i>q</i>	\widehat{VR}_k	Variance ratio				Box–Pierce		Fama–French
		Pooled	Median	Max	Min	Pooled	Max	$t_{\widehat{\beta}_{n,k}}$
Panel A. German mark excess returns: <i>k</i> = 13								
2	1.11	0.06	0.03	0.17	0.02	0.21	0.41	0.07
4	1.35	0.02	0.01	0.05	0.00	0.12	0.14	0.03
8	1.73	0.01	0.00	0.03	0.00	0.37	0.15	0.36
12	1.88	0.02	0.01	0.04	0.00	0.61	0.31	0.52
16	1.91	0.04	0.01	0.08	0.00	0.55	0.19	0.75
20	1.96	0.05	0.02	0.10	0.01	0.67	0.37	0.68
32	1.33	0.24	0.19	0.43	0.08	0.70	0.65	1.00
40	1.10	0.34	0.31	0.58	0.12	0.71	0.65	0.43
Panel B. British pound excess returns: <i>k</i> = 13								
2	1.11	0.08	0.06	0.06	0.13	0.25	0.13	0.10
4	1.22	0.11	0.07	0.12	0.05	0.58	0.06	0.22
8	1.36	0.12	0.10	0.09	0.09	0.48	0.03	0.34
12	1.49	0.11	0.08	0.07	0.08	0.62	0.05	0.70
16	1.50	0.14	0.10	0.10	0.11	0.60	0.06	0.82
20	1.31	0.24	0.20	0.19	0.19	0.37	0.01	0.87
32	0.79	0.50	0.52	0.53	0.52	0.36	0.02	0.86
40	0.62	0.57	0.64	0.65	0.62	0.41	0.02	0.24
Panel C. Japanese yen excess returns: <i>k</i> = 13								
2	1.17	0.01	0.01	0.03	0.00	0.06	0.08	0.01
4	1.31	0.04	0.03	0.02	0.02	0.00	0.06	0.19
8	1.51	0.06	0.03	0.04	0.03	0.01	0.17	0.58
12	1.55	0.10	0.06	0.08	0.05	0.03	0.48	0.77
16	1.42	0.18	0.14	0.17	0.12	0.01	0.30	0.57
20	1.24	0.28	0.25	0.30	0.20	0.02	0.12	0.33
32	1.13	0.34	0.32	0.39	0.26	0.60	0.17	0.55
40	1.33	0.26	0.24	0.29	0.18	0.39	0.26	0.40
<i>T</i>	1716	1716	1716	1716	1716	1716	1716	1716

Notes: This table displays the right tail *p*-values of the tests obtained from the parametric empirical bootstrap distributions. \widehat{VR}_k denotes the estimated pooled variance ratios defined in Eq. (9). See also Notes in Table 1.

12, 16, 20, 32, and 40 quarters are 2, 0, 0, 0, 1, 8, and 12% for the three-month German mark excess return against the U.S. dollar.

As predicted from our Monte Carlo simulations, the Fama–French regression test is not able to reject the null for most values of *q*. On the other hand, the Box–Pierce pooled and maximum *Q* statistics reject the null for the British pound and Japanese yen excess returns. The *p*-values of these statistics tend to decrease and then to increase over *q*. These results further confirm the conclusions from the variance ratio tests.

7. Concluding remarks

This paper investigates both the asymptotic and the finite sample properties of the serial dependence tests for *m*-dependent data. We propose a general econometric framework that first splits the original sample into *m* + 1 subsamples and then aggregates information across them. These aggregation methods include the pooled, median, maximum, minimum, and Wald methods. Our Monte Carlo simulations show that all of these methods except for the Wald perform similarly in terms of size and power.

Using these methods, we conduct tests for the expectations hypothesis in foreign exchange markets and confirm the empirical evidence from previous studies. In addition, we show that the serial dependence tests further provide information regarding the rejections of the null hypothesis. Evidence on positive serial dependence of foreign excess returns supports an expectational error alternative.

Acknowledgments

The research support from Spanish Secretary of Education (SEJ2007-63098) is gratefully acknowledged by the first author.

The research support from the Spanish Plan Nacional de I + D + I (SEJ2007-62908) is gratefully acknowledged by the second author.

Appendix A. Proofs

Proof of Lemma 2.2. The central limit theorem for $U_k(q)$ follows by the Cramer–Wold device and by a standard central limit for *m*-dependent processes, e.g. Diananda (1955), since linear combinations of the cross products involved in autocovariances for different subsamples are also finite dependent for fixed lags. For the asymptotic variance calculation, consider $\widetilde{VR}_k^{(a)}(q) = 1 + 2 \sum_{i=1}^{q-1} \left(1 - \frac{i}{q}\right) \widetilde{\gamma}_k^{(a)}(i) / \sigma_k^2$ with $\widetilde{\gamma}_k^{(a)}(i)$ being the lag *i* sample autocovariance of $\widetilde{\xi}_t = \xi_{t|k} - \alpha_k$ in subsample *a* which centers $\xi_{t|k}$ around the true mean α_k . Then, both $\widehat{VR}_k^{(a)}(q) - 1$ and $\widetilde{VR}_k^{(a)}(q) - 1$ have the same asymptotic distribution since $\widetilde{\gamma}_k^{(a)}(0) \rightarrow_p \sigma_k^2$ for all $a = 1, \dots, k$ and $\widehat{\alpha}_k = \alpha_k + O_p(T^{-1/2})$ under the null. Therefore, from direct calculation from the definition of $\widetilde{VR}_k^{(a)}(q) - 1$, it holds under the null that, for $1 \leq b \leq a \leq k$,

$$\frac{2(q-1)(2q-1)\sigma_k^4}{12q} \sum_k(q)^{[a,b]} = \sum_{i=1}^{q-1} \left(1 - \frac{i}{q}\right)^2 \delta_k^{(a,b)}(i, i) + \sum_{i=1}^{q-1} \sum_{j \neq i} \left(1 - \frac{i}{q}\right) \times \left(1 - \frac{j}{q}\right) \delta_k^{(a,b)}(i, j), \tag{37}$$

where $\delta_k^{(a,b)}(i, j)$ is the asymptotic covariance between $\widetilde{\gamma}_k^{(a)}(i)$ and $\widetilde{\gamma}_k^{(b)}(j)$, $i, j > 0$, normalized by $(T/k)^{1/2}$, given by (5) for $j =$

$i \pm 1$, while $\delta_k^{(a,b)}(i, j) = 0$ for $|j - i| > 1$ and $a \neq b$ because in this case the collection $\{\bar{\xi}_{tk}, \bar{\xi}_{rk+a-b}, \bar{\xi}_{(t+i)k}, \bar{\xi}_{(t+j)k+a-b}\}$ always contains one (zero mean) component independent of the other three ones for any $t, r = 1, \dots, T/k$, and $a, b = 1, \dots, k$. Similarly $\delta_k^{(a,a)}(i, j) = 0$ for $i \neq j$, because $\{\bar{\xi}_{tk}, \bar{\xi}_{rk}, \bar{\xi}_{(t+i)k}, \bar{\xi}_{(t+j)k}\}$ always contains independent individual $\bar{\xi}_t$. \square

Proof of Corollary 2.3. It follows as Lemma 2.2 using Theorem 3 in Lo and MacKinlay (1988) to exploit condition (1) for obtaining the asymptotic variance of autocovariances $\Sigma_k(q)$. \square

Proof of Corollary 2.4. Under $H_0^{(k)}$ and Assumption 2.2, using the iid assumption on the innovations of the MA $(k - 1)$ model,

$$\begin{aligned} E[\bar{\xi}_t \bar{\xi}_{t+a-b} \bar{\xi}_{t+ik} \bar{\xi}_{t+ik+a-b}] &= E[\bar{\xi}_t \bar{\xi}_{t+a-b}] E[\bar{\xi}_{t+ik} \bar{\xi}_{t+ik+a-b}] \\ &= E[\bar{\xi}_0 \bar{\xi}_{a-b}]^2 \\ E[\bar{\xi}_t \bar{\xi}_{t+a-b-k} \bar{\xi}_{t+ik} \bar{\xi}_{t+a-b-k+ik}] &= E[\bar{\xi}_t \bar{\xi}_{t+a-b-k}] \\ &\quad \times E[\bar{\xi}_{t+ik} \bar{\xi}_{t+a-b-k+ik}] \\ &= E[\bar{\xi}_0 \bar{\xi}_{k-a+b}]^2 \\ E[\bar{\xi}_t \bar{\xi}_{t+a-b-k} \bar{\xi}_{t+ik} \bar{\xi}_{t+ik+a-b}] &= E[\bar{\xi}_t \bar{\xi}_{t+a-b-k}] E[\bar{\xi}_t \bar{\xi}_{t+a-b}] \\ &= E[\bar{\xi}_0 \bar{\xi}_{k-a+b}] E[\bar{\xi}_0 \bar{\xi}_{a-b}] \end{aligned}$$

for $i \geq 1$, while for $i \geq 2$,

$$\begin{aligned} E[\bar{\xi}_t \bar{\xi}_{t+a-b} \bar{\xi}_{t+ik} \bar{\xi}_{t+ik+a-b-k}] &= E[\bar{\xi}_t \bar{\xi}_{t+a-b}] E[\bar{\xi}_{t+ik} \bar{\xi}_{t+ik+a-b-k}] \\ &= E[\bar{\xi}_0 \bar{\xi}_{a-b}] E[\bar{\xi}_0 \bar{\xi}_{k-a+b}], \end{aligned}$$

by $(k - 1)$ -dependence, where the pairwise expectations of $\bar{\xi}_t$ do not depend on i by stationarity. Note that for $i = 1$, the first expectation can be calculated as

$$\begin{aligned} E[\bar{\xi}_t \bar{\xi}_{t+a-b} \bar{\xi}_{t+k} \bar{\xi}_{t+k+a-b}] &= E[\bar{\xi}_t \bar{\xi}_{t+a-b}] E[\bar{\xi}_{t+k} \bar{\xi}_{t+k+a-b}] \\ &\quad + E[\bar{\xi}_t \bar{\xi}_{t+k+a-b}] E[\bar{\xi}_{t+a-b} \bar{\xi}_{t+k}] \\ &\quad + E[\bar{\xi}_t \bar{\xi}_{t+k}] E[\bar{\xi}_{t+a-b} \bar{\xi}_{t+k+a-b}] \\ &\quad + \kappa[\bar{\xi}_t, \bar{\xi}_{t+a-b}, \bar{\xi}_{t+k}, \bar{\xi}_{t+k+a-b}], \end{aligned}$$

where κ indicates joint cumulant. Now $E[\bar{\xi}_t \bar{\xi}_{t+a-b}] E[\bar{\xi}_{t+k} \bar{\xi}_{t+k+a-b}] = E[\bar{\xi}_0 \bar{\xi}_{a-b}]^2$, but $E[\bar{\xi}_t \bar{\xi}_{t+k+a-b}] = E[\bar{\xi}_t \bar{\xi}_{t+k}] = 0$, so the second and third terms are zero. Finally the joint cumulant is also zero because it can be written as a (weighted) sum of joint cumulants of the serially independent innovations e_t . Note that all indexes of e_t cannot be the same because $\bar{\xi}_t$ and $\bar{\xi}_{t+k+a-b}$ are independent.

A similar analysis can be done for $a < b$, so that $\delta_k^{(a,b)}(i, i - 1) = \delta_k^{(b,a)}(j, j + 1)$, $i > 1, j > 0$, only depends on $|a - b|$.

Then, the result follows by simple algebra noting that (37) is equal to

$$\begin{aligned} \delta_k^{(a,b)}(i, i) &\sum_{i=1}^{q-1} \left(1 - \frac{i}{q}\right)^2 + 2\delta_k^{(a,b)}(i, i - 1) \\ &\times \sum_{i=2}^{q-1} \left(1 - \frac{i}{q}\right) \left(1 - \frac{i-1}{q}\right), \end{aligned}$$

because $\delta_k^{(a,b)}$ does not depend on i , $q^2 \sum_{i=1}^{q-1} \left(1 - \frac{i}{q}\right)^2 = (q - 1)q(2q - 1)/6$ for $q > 1$, and $q^2 \sum_{i=2}^{q-1} \left(1 - \frac{i}{q}\right) \left(1 - \frac{i-1}{q}\right) = (q - 1)q(2q - 4)/6$ for $q > 2$. Finally, for $s > 0$,

$$E[\bar{\xi}_0 \bar{\xi}_s]^2 = \sum_{i=1}^k \sum_{j=1}^k c_i c_j E(e_{i-k} e_{s+j-k}) = \sigma^2 \sum_{i=1}^{k-s} c_i c_{i-s},$$

since the expectation is only different from zero for $j = i - s$, and the sum is zero when $k - s < 1$, i.e. $s \geq k$. The fact that

$\Sigma_k(q)$ is positive definite follows from stationarity and the non perfect correlation among subsamples under m -dependence. \square

Proof of Lemma 2.5. It follows from direct calculation from Lemma 2.2, where $i = a - b > 0$, and noting that the asymptotic variance of $\sqrt{T} \widehat{VR}_k(q)$ is equal to that of $\frac{1}{k} \sum_{j=1}^k \sqrt{T} \widehat{VR}_k^{(j)}(q)$ because it is not affected by standardization using different, but consistent, estimates of σ_k^2 . Then, $\Lambda_k(q) > 0$ follows from $\Sigma_k(q) > 0$, cf. Lemma 2.2, and Bartlett's weights $(k - i)/k$ in the definition of $\Lambda_k(q)$. \square

Proof of Corollary 2.6. It follows from Lemma 2.5 and Corollary 2.3. \square

Proof of Corollary 2.7. It follows from Lemma 2.5 and Corollary 2.4. \square

Proof of Lemma 2.9. The proof is similar to that of Lemmas 2.2 and 2.5, noting that \hat{y}_k is asymptotically normal and using directly the variance structure of $\hat{y}_k^{(a)}(i)$ to derive $\Xi_k^{(i,j)}(q)$. \square

Proof of Lemma 3.1. We show that the bootstrap versions of VR statistics have the same asymptotic distribution conditional on the sample, as the original VR statistics. Since the bootstrapped residuals \tilde{e}_t^* are iid conditional on the sample, and the estimates \hat{c}_i of c_i are a.s. consistent under $H_0^{(k)}$ and Assumption 2.2, see Yao and Brockwell (2006), the bootstrapped $\tilde{\xi}_t^*$ follow an MA $(k - 1)$ model (up to a negligible initial condition, see e.g. Eq. (2.1) in Bose (1990)). Then, to apply a central limit theorem to $\widehat{VR}_k^{(j)*}(q)$, $j = 1, \dots, k$, or to any finite collection of subsample autocovariances built with this particular $(k - 1)$ -dependent series, we only need to check the moment condition $E^*[\tilde{\xi}_t^{*4}] = O(1)$ a.s., where E^* denotes expectation conditional on the sample. For that we note that $E^*[(\tilde{\xi}_t^*)^4] = E^*[(y_{T+t}^*)^4]$, where

$$\tilde{\xi}_t^* = y_{T+t}^* = \sum_{i=1}^k \hat{c}_i \tilde{e}_{T+t+i-k}^*$$

and $\hat{c}_k = 1$ for identification. Then, we have that

$$E^*[\tilde{\xi}_t^{*4}] = \sum_{i_1=1}^k \dots \sum_{i_4=1}^k \hat{c}_{i_1} \dots \hat{c}_{i_4} E^*[\tilde{e}_{T+t+i_1-k}^* \dots \tilde{e}_{T+t+i_4-k}^*],$$

and $E^*[\tilde{e}_{T+t+i_1-k}^* \dots \tilde{e}_{T+t+i_4-k}^*] \leq E^*[(\tilde{e}_t^*)^4] = T^{-1} \sum_{t=1}^T \tilde{e}_t^4$, with $\tilde{e}_t = \hat{e}_t - \bar{e}_T$. Finally,

$$\hat{e}_t = \xi_t - \sum_{i=1}^{t-1} \hat{\beta}_i \xi_{t-i},$$

where $\hat{\beta}_i$ are the coefficients of $\hat{\beta}(L) = 1/(1 + \hat{c}_{k-1}L + \dots + \hat{c}_1 L^{k-1})$. For large T , $|\hat{\beta}_i| \leq C \delta_k^i$ a.s. for some $0 < \delta_k < 1$. Then, $T^{-1} \sum_{t=1}^T (\tilde{e}_t^*)^4 \rightarrow E[e_t^4] < \infty$ a.s., and the moment condition holds. The covariance structure of subsample bootstrap autocovariances could be checked using the same methods, see, e.g., Lemma 3.1. in Bose (1990), and the lemma follows. \square

Appendix B. Derivations for Section 4

Derivation of Eq. (35).

The “no-bubbles” solution to Eq. (34) is

$$s_t = (1 - \bar{b}) \sum_{i=0}^{\infty} \bar{b}^i E_t w_{t+i*k} + \bar{b} \sum_{i=0}^{\infty} \bar{b}^i E_t \bar{p}_{t+i*k|k}^e \quad (38)$$

Using Eqs. (25), (32) and (38), the expression for the spot exchange rate is obtained

$$s_t = w_t + \frac{\bar{b}}{1 - \bar{b}\psi_k} \bar{p}_{t|k}^e \quad (39)$$

The forward exchange rate is derived from Eqs. (25), (30)–(33) and (39)

$$f_{t|k} = w_t + \left(\frac{\bar{b}\psi_k}{1 - \bar{b}\psi_k} \right) \bar{p}_{t|k}^e + p_{t|k}^e, \quad (40)$$

where $p_{t|k}^e = -\left(\frac{\lambda\bar{b}(\psi_k-1)}{1-\bar{b}\psi_k}\right)\bar{p}_{t|k}^e - \lambda g \varpi_t$. Finally, we derive the foreign excess return in (35) by combining (39) with (40). \square

Derivation of Eq. (36).

Assuming that PPP holds and $E_t^r[s_{t+k}] = E_{t-k}[s_t]$, we have a setup for s_t ,

$$s_t = bE_t[s_{t+k}] + bp_{t|k}^e + (1 - b)w_t, \quad (41)$$

where $p_{t|k}^e = E_t^m[s_{t+k}] - E_t[s_{t+k}] = -\lambda(E_t[s_{t+k}] - E_{t-k}[s_t])$. By taking conditional expectations on the information set at $t - k$ in both sides of Eq. (41) and solving forward, we have

$$E_{t-k}[s_t] = (1 - \bar{b}) \sum_{i=0}^{\infty} \bar{b}^i E_{t-k} w_{t+i*k}. \quad (42)$$

From (25) and (42), we have $E_{t-k}[s_t] = w_{t-k}$. Inserting this into (41), we have

$$s_t = (1 - b\lambda)w_t + b\lambda w_{t-k}. \quad (43)$$

Here, s_t is a weighted average of w_t and w_{t-k} where the weight is determined by λ . And the forward exchange rate is

$$f_{t|k} = E_t^m[s_{t+k}] = (1 - \lambda)w_t + \lambda w_{t-k}. \quad (44)$$

Finally, we derive the foreign excess return in (36) by using (43) and (44). \square

References

- Baillie, R., Bollerslev, T., 2000. The forward premium anomaly is not as bad as you think. *Journal of International Money and Finance* 19, 471–488.
- Bose, A., 1990. Bootstrap in moving average models. *Annals of the Institute of Statistical Mathematics* 42, 753–768.
- Box, G., Pierce, D., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–1526.
- Bühlmann, P., 1997. Sieve bootstrap for time series. *Bernoulli* 3, 123–148.
- Campbell, B., Dufour, J., 1995. Exact nonparametric orthogonality and random walk tests. *The Review of Economics and Statistics* 77, 1–16.
- Campbell, J., Lo, A., MacKinlay, A., 1997. *The Econometrics of Financial Markets*. Princeton University Press, Princeton.
- Campbell, J., Shiller, R., 1991. Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies* 58, 495–514.
- Campbell, J., Yogo, M., 2006. Efficient tests of stock return predictability. *Journal of Financial Economics* 81, 27–60.
- Cochrane, J., 1988. How big is the random walk in GNP? *Journal of Political Economy* 96, 893–920.
- Cochrane, J., Piazzesi, M., 2005. Bond risk premia. *American Economic Review* 95, 138–160.
- Corradi, V., Iglesias, E., 2008. Bootstrap refinements for QML estimators of the GARCH(1, 1) parameters. *Journal of Econometrics* 144, 500–510.
- Delgado, M., Velasco, C., 2011. An asymptotically pivotal transform of the residuals sample autocorrelations with application to model checking. *Journal of the American Statistical Association* 106, 946–958.
- Diananda, P., 1955. The central limit theorem for m -dependent variables. *Mathematical Proceedings of the Cambridge Philosophical Society* 51, 92–95.
- Dufour, J., Torres, O., 1998. Union-intersection and sample-split methods in econometrics with applications to SURE and MA models. In: Giles, D., Ullah, A. (Eds.), *Handbook of Applied Economic Statistics*. Marcel Dekker, New York, Chapter 14.
- Engel, C., 1996. The forward discount anomaly and the risk premium: A survey of recent evidence. *Journal of Empirical Finance* 3, 123–192.
- Engel, C., West, K., 2005. Exchange rates and fundamentals. *Journal of Political Economy* 113, 485–517.
- Escanciano, J., Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics* 134, 151–185.
- Fama, E., French, K., 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96, 246–273.
- Frankel, J., Froot, K., 1987. Using survey data to test standard propositions regarding exchange rate expectations. *American Economic Review* 77, 133–153.
- Frankel, J., Froot, K., 1990. Chartists, fundamentalists, and demand for dollars. In: Courakis, Anthony S., Taylor, Mark P. (Eds.), *Private Behavior and Interdependent Government Policy*. Clarendon Press, Oxford.
- Giacomini, R., Politis, N., White, H., 2012. A warp-speed method for conducting monte carlo experiments involving bootstrap estimators. *Mimeo*.
- Godfrey, L., 1978. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica* 46, 1293–1301.
- Hansen, L., Hodrick, R., 1980. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy* 88, 829–853.
- Hidalgo, J., Zaffaroni, P., 2007. A goodness-of-fit test for ARCH(∞) model. *Journal of Econometrics* 141, 835–875.
- Jansson, M., Moreira, M., 2006. Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74, 681–714.
- Kreiss, J-P., Franke, J., 1992. Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis* 13, 297–321.
- Künsch, H., 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
- Lewis, K., 1995. Puzzles in international financial markets. In: Grossman, Gene, Rogoff, Kenneth (Eds.), *Handbook of International Economics Vol. 3*. North Holland, Amsterdam.
- Liu, C., He, J., 1991. A variance-ratio test of random walks in foreign exchange rates. *Journal of Finance* 46, 773–785.
- Ljung, G., Box, G., 1978. On a measure of a lack of fit in time series models. *Biometrika* 65, 297–303.
- Lo, A., MacKinlay, C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1, 41–66.
- Lo, A., MacKinlay, C., 1989. The size and power of the variance ratio tests in finite samples: A monte carlo investigation. *Journal of Econometrics* 40, 203–238.
- Moon, S., Velasco, C., 2011. The forward discount puzzle: identification of economic assumptions. *Universidad Carlos III de Madrid*.
- Newey, W., West, K., 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Obstfeld, M., Rogoff, K., 2002. Risk and exchange rates. In: Helpman, Elhanan, Sadka, Efraim (Eds.), *Contemporary Economic Policy: Essays in Honor of Assaf Razin*. Cambridge University Press, Cambridge.
- Poterba, J., Summers, L., 1988. Mean reversion in stock prices: evidence and implications. *Journal of Financial Economics* 22, 27–59.
- Richardson, M., Smith, T., 1991. Tests of financial models in the presence of overlapping observations. *The Review of Financial Studies* 4, 227–254.
- Richardson, M., Stock, J., 1989. Drawing inferences from statistics based on multiyear asset returns. *Journal of Financial Economics* 25, 323–348.
- Tauchen, G., 2001. The bias of tests for a risk premium in forward exchange rates. *Journal of Empirical Finance* 8, 695–704.
- Wright, J., 2000. Alternative variance ratio tests using ranks and signs. *Journal of Business and Economic Statistics* 18, 1–9.
- Yao, Q., Brockwell, P., 2006. Gaussian maximum likelihood estimation for ARMA models I Time Series. *Journal of Times Series Analysis* 27, 857–875.