Comment

Matias D. Cattaneo
Department of Economics, University of Michigan, Ann Arbor, MI 48109 (cattaneo@umich.edu)

Richard K. Crump
Capital Markets Function, Federal Reserve Bank of New York, New York, NY 10045 (richard.crump@ny.frb.org)

1. INTRODUCTION

Conducting valid inference in a time series setting often requires the use of a heteroscedasticity and autocorrelation (HAC) robust variance estimator. Under mild dependence structures, the theory and practice of such estimators is well developed. However, under more severe forms of dependence, the conventional distributional approximation usually employed to describe the finite-sample properties of test statistics based on these estimators tends to be poor. Professor Müller is to be congratulated for this excellent article addressing the important issue of conducting valid inference using HAC estimators in the presence of strong autocorrelation. The class of tests introduced in the article should prove useful both to applied practitioners and as a foundation for future theoretical work.

This comment is comprised of two sections. First, we compare the main contribution of Müller (2014) (hereafter, the "S_q test") to a theoretically valid approach based on the usual t-test. More specifically, to gain further insight into the properties of the S_q test, we compare its finite-sample properties to those of a t-test with limiting distribution obtained under the local-to-unity parameterization and fixed-b asymptotics. We use critical values obtained from a Bonferroni-based procedure to control size in the presence of the nuisance parameter governing the degree of dependence. Second, we employ the new test in an empirical application by revisiting the question of long-horizon predictability in asset returns. We find that the S_q test provides evidence of predictability of equity returns by the dividend yield at shorter horizons when the sample is restricted to end in 1990. The S_q test, when applied to bond returns, produces little evidence of predictability in our application. In both applications the conclusions drawn from the S_q test can be sensitive to the choice of q, suggesting that further work will be necessary to guide the use of the S_q test in empirical applications.

2. FIXED-b ASYMPTOTICS IN A LOCAL-TO-UNITY SETTING

The new testing procedure of Müller (2014) is motivated by the assumption that for a restricted class of frequencies, governed by the user-defined parameter q, the spectral density is well approximated by the spectral density of a nearly integrated autoregressive process. By focusing only on this band of frequencies, the core assumption of the article is of the "semiparametric" variety. The exact form of the S_q test is then derived under the assumption of scale invariance and maximization of a weighted-average power criterion using the results of Elliott, Müller, and Watson (2013).

Suppose that instead of making the semiparametric assumption of Müller (2014), we assume $\{y_t : t = 1, \ldots, T\}$ is generated by

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \tag{2.1}$$

with initial condition ε_0, where $\{\eta_t\}$ is a weakly dependent process. Furthermore, we impose the local-to-unity parameterization, $\rho_T = 1 - c/T$, where $c \in [0, \infty)$. In words, we assume that the data are generated by a nearly integrated autoregressive process at all frequencies. As in Müller (2014), we are interested in testing $\mathbb{H}_0 : \mu = \mu_0$ versus the alternative that $\mathbb{H}_A : \mu \neq \mu_0$. We use the fixed-$b$ asymptotics of Kiefer and Vogelsang (2005) in this setting. Following, for example, Atchadé and Cattaneo (2014), we can write the long-run variance estimator as

$$\hat{\omega}^2_{k,S_T} = T^{-3}\sum_{\ell=1}^{T}\sum_{j=1}^{T}\left\{k_b\left(\frac{\ell-j}{S_T}\right) - \upsilon_T(\ell) - \upsilon_T(j) + u_T\right\} \times y_\ell y_j$$

$$\upsilon_T(\ell) = T^{-1}\sum_{i=1}^{T} k_b\left(\frac{\ell-i}{S_T}\right), \quad u_T = T^{-2}\sum_{i=1}^{T}\sum_{j=1}^{T} k_b\left(\frac{j-i}{S_T}\right),$$

where $k_b(\cdot)$ is a kernel function. In the simulations, we use a Parzen kernel, $b = 0.5$ or $b = 1$, and $S_T = T$. Then, under regularity conditions, and defining $\hat{\mu} = T^{-1}\sum_{t=1}^{T} y_t$, we have

$$\tau_b = \frac{T^{-1/2}(\hat{\mu}-\mu)}{\sqrt{\hat{\omega}^2_{k,S_T}}} \xrightarrow{d} \frac{\int_0^1 B_c(r)dr}{\int_0^1\int_0^1 K_b(s,t)B_c(s)B_c(t)dsdt},$$

$$K_b(s,t) = k_b(s-t) - \int_0^1 k_b(s-w)\,dw - \int_0^1 k_b(t-w)\,dw + \int_0^1\int_0^1 k_b(w_1-w_2)\,dw_1dw_2,$$

where $\{B_c(s) : s \in [0, 1]\}$ is an Ornstein–Uhlenbeck process. For instance, this result may be obtained from similar steps as in Tanaka (1996, chap. 5). The limiting distribution of the t-test, τ_b, is then solely a function of the user-defined choice of kernel

Journal of Business & Economic Statistics
July 2014, Vol. 32, No. 3
DOI: 10.1080/07350015.2014.928220

Table 1. AR(1), $\varepsilon_0 \sim \mathcal{N}\left(0, \sigma_\eta^2/(1-\rho^2)\right)$

ρ	S_{12}	S_{24}	S_{48}	$\tau_{1/2}$	τ_1	$\tau^\star_{1/2}$	$\tau^\star_1$
				Panel A: size			
0	4.9	4.7	4.7	5.5	5.3	5.5	5.7
0.7	5.0	4.9	4.8	5.2	5.1	5.2	5.5
0.9	5.0	5.0	5.2	4.1	3.9	4.2	4.3
0.95	5.0	5.1	5.2	3.5	3.4	3.5	3.3
0.98	4.9	4.9	5.1	5.0	5.4	3.1	2.9
0.999	4.4	4.4	4.2	48.0	41.6	5.2	4.9
				Panel B: power			
0	34.2	42.1	47.1	36.5	27.8	36.3	29.2
0.7	33.8	40.6	44.3	34.7	26.7	34.9	28.1
0.9	28.0	32.9	35.5	26.1	21.7	26.6	23.0
0.95	20.2	22.5	24.7	19.0	18.8	16.7	14.8
0.98	11.2	11.7	12.0	29.8	27.7	10.1	9.3
0.999	4.9	5.0	4.9	100.0	100.0	8.1	8.0
				Panel C: size-adjusted power			
0	34.6	43.1	47.9	34.0	26.7	34.0	26.7
0.7	33.6	40.9	45.1	34.6	26.2	34.6	26.2
0.9	27.8	33.0	34.3	34.6	27.8	34.6	27.8
0.95	20.1	22.1	23.4	34.3	28.2	34.3	28.2
0.98	11.4	12.2	11.7	36.1	27.9	36.1	27.9
0.999	5.6	5.8	5.8	92.1	80.4	92.1	80.4

and parameter b and the nuisance parameter c, which governs the degree of persistence of the data. To conduct valid inference, it is crucial to control the size of the test with respect to the value of c. See, for example, Andrews and Guggenberger (2009) and Andrews and Guggenberger (2010) for further discussion on the importance of uniformly valid inference in econometrics. We use the Bonferroni-based critical values introduced in McCloskey (2012). We view this approach as possibly the most natural point of comparison to the S_q test.

Table 1 presents the simulation results for the strictly stationary AR(1) model and set of alternatives given in Müller (2014). As in Müller (2014), we set $\sigma_\eta^2 = 1$. The first three columns present the size, power, and size-adjusted power (we show size-adjusted power so the results for the S_q test are comparable to the tables in Müller (2014). Size-adjusted power for the τ_b statistic is (nearly) the same as inference when the value of c is known and so should be interpreted with this in mind) for the S_q test with $q = 12$, $q = 24$, and $q = 48$. The next two columns provide the results for the Bonferroni-based procedure with critical values formed under the assumption that the initial condition is negligible (written as τ_b). As an additional point of comparison, the final two columns report results for the Bonferroni-based procedure with the limiting distributions constructed under a stationary initial condition (since $c > 0$ in this case, we implemented the testing procedure by setting the lower and upper critical values for μ equal to $-\infty$ and ∞, respectively, whenever the confidence interval formed for c included the smallest value in our grid. This approach thus requires a user-defined minimum value of c. In our simulations, we chose $c_{\min} = 0.01$) (written as $\tau_b^\star$). The "S-Bonf-Adj" critical values are formed with $\beta = 0.15$ (i.e., nominal coverage of the parameter c equal to $1-\beta$). To form confidence intervals for c, we invert an ordinary least square (OLS)-based augmented Dickey–Fuller (ADF) test with lag length chosen by the modified Akaike information criterion (MAIC) of Perron and Qu (2007). Refinements of our procedure could include shifting to the "S-Bonf-Min" critical values of McCloskey (2012), a different choice of confidence interval for the local-to-unity parameter or an alternative test statistic to τ. However, we prefer this formulation for simplicity of interpretation. All results are based on 20,000 simulations.

The results of Table 1 are instructive on how to interpret properties of the S_q test. First, the τ_b test statistic controls size away from values of ρ close to one, but is severely size distorted when $\rho = 0.999$. This reflects the fact that the critical values for τ_b are constructed under the assumption that the initial condition is negligible. Meanwhile, the $\tau_b^\star$ test statistic controls size well across this grid of values of ρ. The power of the $\tau^\star_{1/2}$ test is comparable to that of the S_{12} test. However, the S_{24} and S_{48} tests have higher power when ρ moves away from one. In Table 2, we again consider the AR(1) specification but use the fixed initial condition $\varepsilon_0 = 0$. The pattern of the results is similar to that of Table 1 except that the τ_b test statistic controls size even in cases where the error terms are highly persistent. We can also contrast the results from Tables 1 and 2 to those using the least-favorable critical value. In this setting, because critical values for τ_b grow as $c \downarrow 0$, choosing to use the least-favorable critical value produces severely undersized tests for all areas of the parameter space except when ρ is very close to one.

In Table 3, we present results for the "AR(1) + noise" specification and set of alternatives given in Müller (2014). As in Müller (2014), we set the variance of the additive noise term to 4. Similar to Table 1, the τ_b test is oversized when ρ is near one with excessive size distortion when $\rho = 0.999$. In contrast, the $\tau_b^\star$ test controls size well across the grid of ρ values and the $\tau^\star_{1/2}$

Table 2. AR(1), $\varepsilon_0 = 0$

ρ	S_{12}	S_{24}	S_{48}	$\tau_{1/2}$	τ_1	$\tau^\star_{1/2}$	$\tau^\star_1$
			Panel A: size				
0	4.8	4.8	4.8	5.6	5.4	5.5	5.7
0.7	5.0	4.9	4.8	5.3	5.1	5.3	5.5
0.9	5.3	5.1	5.5	4.2	4.0	4.4	4.5
0.95	5.4	5.4	5.6	3.6	3.5	3.5	3.4
0.98	5.0	4.8	5.0	4.1	4.7	3.3	3.0
0.999	2.5	2.4	2.3	6.0	5.3	3.0	2.5
			Panel B: power				
0	34.5	42.5	47.3	36.7	28.0	36.4	29.3
0.7	34.3	41.3	44.9	35.2	27.2	35.4	28.5
0.9	29.7	35.1	37.7	27.1	22.4	27.6	23.9
0.95	22.9	25.7	28.1	20.1	20.1	17.6	15.9
0.98	13.8	14.9	15.6	34.0	32.5	11.0	10.4
0.999	5.2	5.3	5.1	100.0	100.0	8.3	8.2
			Panel C: size-adjusted power				
0	35.0	43.1	48.1	34.3	27.5	34.3	27.5
0.7	34.4	41.6	45.8	34.9	27.4	34.9	27.4
0.9	28.9	34.5	35.5	35.7	28.7	35.7	28.7
0.95	21.6	23.9	25.4	37.7	29.9	37.7	29.9
0.98	13.8	15.9	15.4	48.7	38.1	48.7	38.1
0.999	10.2	12.6	11.1	100.0	100.0	100.0	100.0

statistic has comparable power properties to that of the S_{12} test. The S_{24} and S_{48} tests suffer from size distortion for larger values of ρ as discussed in Müller (2014). Table 4 reports the companion results under a zero initial condition. As in Table 3, the S_{12} test and the $\tau^\star_b$ control size well across these values of ρ. The S_{24} and S_{48} tests show some size distortion as ρ moves above 0.90 whereas the τ_b test controls size except when $\rho = 0.999$.

There are three main observations from the limited simulation evidence we present. First, it is instructive to see where power is directed by the S_q test. In particular, the S_q test has little power when ρ is near 1 but much higher power elsewhere. From a practitioner's perspective, this is a very appealing property (see next section) as the test controls size well in exactly the region of the parameter space where

Table 3. AR(1) + Noise, $\varepsilon_0 \sim \mathcal{N}\left(0, \sigma^2_\eta/(1-\rho^2)\right)$

ρ	S_{12}	S_{24}	S_{48}	$\tau_{1/2}$	τ_1	$\tau^\star_{1/2}$	$\tau^\star_1$
			Panel A: size				
0	4.9	5.0	5.0	5.5	5.3	5.4	5.5
0.7	4.9	4.9	5.6	5.2	4.9	5.2	5.1
0.9	5.1	5.9	9.2	4.9	4.2	5.0	4.3
0.95	5.2	6.7	12.4	4.7	4.3	4.7	4.0
0.98	5.3	7.0	15.0	5.9	6.0	4.0	3.2
0.999	4.9	7.4	17.0	48.1	41.9	4.8	4.5
			Panel B: power				
0	34.4	42.7	47.2	35.5	27.1	35.1	27.9
0.7	34.1	42.5	49.1	33.9	26.3	33.7	26.8
0.9	28.7	37.9	53.7	27.1	22.8	26.7	21.6
0.95	22.0	29.8	50.0	21.2	20.7	17.6	14.7
0.98	12.4	17.8	37.3	30.9	28.7	9.6	8.6
0.999	5.5	8.4	19.4	100.0	100.0	6.8	6.7
			Panel C: size-adjusted power				
0	34.8	42.7	47.3	34.4	26.2	34.4	26.2
0.7	34.6	42.9	47.0	36.2	26.9	36.2	26.9
0.9	28.3	34.1	38.0	35.7	29.1	35.7	29.1
0.95	21.2	23.0	25.4	35.0	28.2	35.0	28.2
0.98	11.5	11.9	12.0	36.6	28.4	36.6	28.4
0.999	5.7	5.7	5.6	92.4	80.6	92.4	80.6

Table 4. AR(1) + Noise, $\varepsilon_0 = 0$

ρ	S_{12}	S_{24}	S_{48}	$\tau_{1/2}$	τ_1	$\tau^{\star}_{1/2}$	$\tau^{\star}_1$
Panel A: size							
0	4.9	4.9	5.0	5.5	5.3	5.4	5.6
0.7	4.9	4.9	5.5	5.2	4.9	5.1	5.1
0.9	5.3	6.0	9.6	4.9	4.5	5.0	4.6
0.95	5.6	7.1	12.6	4.7	4.4	4.6	4.1
0.98	5.4	6.9	14.7	5.4	5.5	4.5	3.6
0.999	2.9	4.0	9.8	7.1	6.3	3.7	2.8
Panel B: power							
0	34.5	42.8	47.3	35.5	27.2	35.2	27.9
0.7	34.5	43.0	49.5	34.2	26.6	34.0	27.1
0.9	30.5	40.1	55.4	28.2	24.0	27.6	22.6
0.95	24.7	33.6	54.4	22.5	22.3	18.5	15.8
0.98	15.4	22.4	44.5	35.0	33.6	10.2	9.6
0.999	5.9	8.7	20.3	100.0	100.0	6.9	6.8
Panel C: size-adjusted power							
0	34.8	43.1	47.2	34.4	26.1	34.4	26.1
0.7	34.9	43.2	47.0	36.7	27.4	36.7	27.4
0.9	29.3	35.9	39.6	36.4	28.4	36.4	28.4
0.95	22.8	25.3	27.4	37.9	30.2	37.9	30.2
0.98	14.1	14.6	15.5	47.7	38.3	47.7	38.3
0.999	10.1	11.6	10.7	100.0	100.0	100.0	100.0

inference is most difficult. Second, the results show clearly that the underlying assumption of how the initial condition is generated plays a key role in the performance of testing procedures for large values of ρ. In particular, procedures that may perform well in terms of size control when the initial condition is negligible can have severe size distortion when the initial condition is drawn from its unconditional distribution. This is not a surprising result, but one that may not be fully appreciated outside of the unit root testing literature. Third, the simulations show that the choice of q can matter and so further guidance for applied practitioners would be an important contribution for future work.

3. APPLICATION: LONG-HORIZON RETURN PREDICTABILITY

Conventional wisdom in applied time series would suggest the presence of model misspecification when products of estimated error terms are highly persistent. The consummate example would be the omission of the lagged left-hand side variable in the canonical spurious regression case. However, an important situation where the prescription to add lags of the dependent variable is unavailable is in long-horizon predictability regressions, such as predicting h-period ahead stock and bond returns. We revisit the question of longer-horizon return predictability for equities and bonds using the S_q test. These two simple empirical exercises demonstrate clearly the applicability of the new testing procedure.

Long-horizon stock return predictability has been studied by a number of authors (see, e.g., Koijen and Nieuwerburgh 2011 or Rapach and Zhou 2013 for general discussions.) Here, we follow the standard approach in the literature and form continuously compounded, cumulative stock returns as

$$rx_{t+h} = \sum_{j=1}^{h} rx_{t+j}, \tag{3.1}$$

where rx_t is the 1 month compounded excess return formed as the Center for Research in Security Prices (CRSP) value-weighted return less the 1 month interest rate (here measured as the Fama–Bliss risk-free rate). We consider two regressors: (1) the log dividend yield formed as the natural logarithm of the sum of monthly dividends over the last 12 months relative to the current price and (2) the Fama–Bliss risk-free rate. We include the latter regressor as Ang and Bekaert (2007) argued that the predictive ability of the dividend yield is enhanced by including this variable. We consider two sample periods: monthly data over the period 1952–2012 and over the period 1952–1990. We include the latter, shorter sample, as it is perceived that the dividend yield was a more reliable predictor of stock returns up to 1990. Finally, we report results for values of $h \in \{6, 12, 24, 36\}$ months. Because the returns in Equation (3.1) are calculated with overlapping periods, they have a considerable degree of persistence, comparable to that of the right-hand side variables. Thus, concerns have arisen about conducting inference in such a setting.

Confidence intervals formed using Newey–West (NW) standard errors with lag truncation parameter h and those using Müller (2014) are reported in Table 5. When the sample is restricted to 1952–1990, the S_q test provides evidence of return predictability at shorter horizons, although the confidence intervals vary considerably depending on the specification and the choice of q. However, there is little evidence of predictability

Table 5. Long-horizon regressions: Equity returns

1952–1990				
Dividend yield only				
h	6	12	24	36
$\hat{\beta}_{dy}$	16.93	33.818	52.339	54.685
NW	(7.357, 26.503)	(15.823, 51.814)	(17.213, 87.464)	(14.407, 94.964)
S_{12}	(2.736, 54.803)	(3.619, 118.719)	(−16.271, 203.980)	(−14.234, 221.210)
S_{24}	(3.652, 33.872)	(5.33, 71.531)	(−5.834, 194.429)	(−22.098, 148.099)
S_{48}	(4.663, 32.203)	(5.334, 76.591)	(−29.8, 125.125)	(−29.001, 159.285)
Dividend yield and risk-free rate				
h	6	12	24	36
$\hat{\beta}_{dy}$	26.435	49.697	69.028	71.753
NW	(17.478, 35.392)	(34.268, 65.125)	(43.526, 94.531)	(44.859, 98.648)
S_{12}	(17.908, 429.574)	(32.07, 874.616)	(−247.772, 155.535)	$(-\infty, \infty)$
S_{24}	(15.647, 52.028)	(25.505, 77.308)	(−115.492, 102.134)	(−323.597, 108.84)
S_{48}	$(-\infty, -548.666) \cup (14.642, \infty)$	$(-\infty, -375.173) \cup (22.021, \infty)$	$(-\infty, 97.251) \cup (534.008, \infty)$	$(-\infty, 118.368) \cup (213.433, \infty)$
1952–2012				
Dividend yield only				
h	6	12	24	36
$\hat{\beta}_{dy}$	5.23	10.564	19.103	23.983
NW	(0.791, 9.668)	(1.268, 19.859)	(1.699, 36.508)	(4.152, 43.815)
S_{12}	$(-\infty, -10.393) \cup (-2.058, \infty)$	$(-\infty, -19.806) \cup (-4.435, \infty)$	$(-\infty, -28.866) \cup (-10.372, \infty)$	$(-\infty, \infty)$
S_{24}	$(-\infty, -22.479) \cup (-3.448, \infty)$	$(-\infty, -47.079) \cup (-6.829, \infty)$	$(-\infty, -63.666) \cup (-16.849, \infty)$	$(-\infty, \infty)$
S_{48}	$(-\infty, -33.192) \cup (-3.911, 14.364) \cup (30.886, \infty)$	$(-\infty, -67.765) \cup (-10.525, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
Dividend yield and risk-free rate				
h	6	12	24	36
$\hat{\beta}_{dy}$	8.397	15.840	25.917	30.976
NW	(3.924, 12.871)	(6.101, 25.580)	(7.795, 44.038)	(11.943, 50.010)
S_{12}	$(-\infty, -9.893) \cup (2.884, \infty)$	$(-\infty, -14.318) \cup (4.238, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
S_{24}	$(-\infty, -16.464) \cup (-1.34, \infty)$	$(-\infty, -31.862) \cup (-5.298, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
S_{48}	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$

NOTE: This table shows the results for long-horizon predictability regressions of equity market returns on the log dividend yield and the risk-free rate. $\hat{\beta}_{dy}$ is the OLS coefficient corresponding to the log dividend yield. Newey–West (NW) standard errors are constructed with h lags. All confidence intervals have nominal coverage of 95%.

at shorter horizons for the full sample. At longer horizons there is no evidence of predictability for either the restricted or full sample based on the S_q test.

Next, we consider a similar exercise for excess bond returns. Specifically, we revisit the influential work of Cochrane and Piazzesi (2005, 2008), where the authors form a bond-return forecasting factor using linear combinations of forward rates. To proceed, we first must form this return-forecasting factor (hereafter, CP factor). We use excess returns and log forward rates, defined by

$$rx_{t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)}, \qquad f_t^{(n)} \equiv p_t^{(n-1)} - p_t^{(n)},$$

where $p_t^{(n)}$ is the log price of an n year discount bond at time t and $y_t^{(1)}$ is the 1-month GSW rate (GSW refers to zero-coupon bond yields from Gurkaynak, Sack, and Wright (2007), which are available at a daily frequency on the Board of Governors of the Federal Reserve's research data page). We use GSW yields to construct excess returns and Fama–Bliss forward rates as regressors. We follow Cochrane and Piazzesi (2008) and regress 14 excess returns $rx_{t+1} = [rx_{t+1}^{(2)}, rx_{t+1}^{(3)}, \ldots, rx_{t+1}^{(15)}]'$ on a constant $z_t = 1$ and five forward rates $w_t = [y_t^{(1)}, f_t^{(2)}, \ldots, f_t^{(5)}]'$. Cochrane and Piazzesi (2008) formed the CP factor by taking the first principal component of the fitted values from this regression. It can be shown that this is equivalent to the maximum-likelihood estimator (MLE) of a reduced-rank regression (with coefficient matrix of rank one) under the assumption of iid Gaussian errors and a scalar variance matrix. We also consider a weighted version of the CP factor, formed as the MLE under the assumption of a diagonal variance matrix (see Adrian, Crump, and Moench 2014 for further details).

Table 6. Long-horizon regressions: Bond returns

	In-sample	Out-of-sample (5 years)	Out-of-sample (10 years)	Out-of-sample (15 years)
Identity weight matrix				
$\hat{\beta}_{\mathrm{CP}}$	0.220	0.052	0.082	0.063
NW	(0.101, 0.339)	(−0.001, 0.105)	(0.047, 0.117)	(−0.009, 0.136)
S_{12}	$(-\infty, 0.328) \cup (3.545, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
S_{24}	(−0.102, 0.324)	(−0.14, 0.197)	$(-\infty, 0.125) \cup (0.225, \infty)$	$(-\infty, -8.129) \cup (-0.031, \infty)$
S_{48}	(−0.106, 0.356)	(−0.165, 0.13)	$(-\infty, -0.277) \cup (-0.036, 0.125) \cup (0.456, \infty)$	(−0.07, 0.741)
Diagonal weight matrix				
$\hat{\beta}_{\mathrm{CP}}$	0.231	0.053	0.086	0.066
NW	(0.105, 0.356)	(−0.003, 0.108)	(0.049, 0.122)	(−0.012, 0.144)
S_{12}	$(-\infty, 0.355) \cup (2.257, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
S_{24}	(−0.116, 0.342)	(−0.155, 0.209)	$(-\infty, 0.132) \cup (0.226, \infty)$	$(-\infty, -13.464) \cup (-0.035, \infty)$
S_{48}	(−0.107, 0.376)	(−0.178, 0.134)	$(-\infty, -0.205) \cup (-0.045, 0.13) \cup (0.419, \infty)$	(−0.078, 0.843)

NOTE: This table shows the results for long-horizon predictability regressions of 1 year excess holding period returns on the CP factor. $\hat{\beta}_{\mathrm{CP}}$ is the OLS coefficient corresponding to the CP factor. The first column report results for the CP factor constructed on data for 1971–2012. The next three columns report results for the CP factor constructed in real time with a 5, 10, and 15 year burn-in period, respectively. Newey–West (NW) standard errors are constructed with 12 lags. All confidence intervals have nominal coverage of 95%.

We then regress average excess returns on the CP factor,

$$\overline{rx}_{t+1} = \alpha + \beta x_t + \epsilon_t,$$

where $\overline{rx}_{t+1}$ is the average return, across maturities, $\overline{rx}_{t+1} = \frac{1}{14}\sum_{n=2}^{15} rx_{t+1}^{(n)}$.

We use the S_q test statistic to construct confidence intervals (results reported in Table 6). We then repeat the exercise but now we construct the CP factor without using future information after a certain burn-in period. (We also considered specifications that added the term spread as an additional predictor. The results in this case were qualitatively similar to those presented here.) We find that the conclusions drawn from the confidence intervals formed from the S_q test are sensitive to different values of q and different burn-in periods. As in the equity application, we find that confidence intervals can be asymmetric, sometimes disjoint but nonempty (as discussed in Müller 2014). Despite this, we find no evidence that a predictive relationship can be uncovered with these data.

ACKNOWLEDGMENTS

The authors thank Lutz Kilian, Adam McCloskey, Emanuel Moench, and Ulrich Müller for helpful comments and discussions and the editors, Keisuke Hirano and Jonathan Wright, for inviting us to participate in this intellectual exchange. Benjamin Mills provided excellent research assistance. The first author gratefully acknowledges financial support from the National Science Foundation (SES 1122994). The views expressed in this article are those of the authors and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System.

REFERENCES

Adrian, T., Crump, R. K., and Moench, E. (2014), "Regression-based Estimation of Dynamic Asset Pricing Models," Staff Report 493, Federal Reserve Bank of New York. [328]

——— (2009), "Hybrid and Size-corrected Subsampling Methods," *Econometrica*, 77, 721–762. [325]

Andrews, D. W. K., and Guggenberger, P. (2010), "Asymptotic Size and a Problem with Subsampling and with the *m* Out of *n* Bootstrap," *Econometric Theory*, 26, 426–468. [325]

Ang, A., and Bekaert, G. (2007), "Stock Return Predictability: Is it There?," *Review of Financial Studies*, 20, 651–707. [327]

Atchadé, Y. F., and Cattaneo, M. D. (2014), "A Martingale Decomposition for Quadratic Forms of Markov Chains (with Applications)," *Stochastic Processes and Their Applications*, 124, 646–677. [324]

Cochrane, J., and Piazzesi, M. (2005), "Bond Risk Premia," *American Economic Review*, 95, 138–160. [328]

——— (2008), "Decomposing the Yield Curve," Working Paper, University of Chicago. [328]

Elliott, G., Müller, U. K., and Watson, M. W. (2013), "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," Working Paper, Princeton University. [324]

Gurkaynak, R. S., Sack, B., and Wright, J. H. (2007), "The U.S. Treasury Yield Curve: 1961 to the Present," *Journal of Monetary Economics*, 54, 2291–2304. [328]

Kiefer, N. M., and Vogelsang, T. J. (2005), "A New Asymptotic Theory for Heteroskedasticity-autocorrelation Robust Tests," *Econometric Theory*, 21, 1130–1164. [324]

Koijen, R. S., and Nieuwerburgh, S. V. (2011), "Predictability of Returns and Cash Flows," *Annual Review of Financial Economics*, 3, 467–491. [327]

McCloskey, A. (2012), "Bonferroni-Based Size-Correction for Nonstandard Testing Problems," Working Paper, Brown University. [325]

Müller, U. K. (2014), "HAC Corrections for Strongly Autocorrelated Time Series," *Journal of Business and Economic Statistics*, 32, 311–322. [324,325,327,329]

Perron, P., and Qu, Z. (2007), "A Simple Modification to Improve the Finite Sample Properties of Ng and Perron's Unit Root Tests," *Economics Letters*, 94, 12–19. [325]

Rapach, D. E., and Zhou, G. (2013), "Forecasting Stock Returns," in *Handbook of Economic Forecasting (Vol. II)*, eds. G. Elliott and A. Timmermann, Amsterdam: North-Holland, pp. 329–383. [327]

Tanaka, K. (1996), *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, New York: Wiley. [324]