

2014s-36

# **Horizon Effect in the Term Structure of Long-Run Risk-Return Trade-Offs**

*Cédric Okou, Éric Jacquier*

---

**Série Scientifique**  
*Scientific Series*

---

**Montréal  
Juillet 2014**

© 2014 Cédric Okou, Éric Jacquier. Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©.  
*Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*



Centre interuniversitaire de recherche en analyse des organisations

## **CIRANO**

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

*CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.*

## **Les partenaires du CIRANO**

### **Partenaire majeur**

Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie

### **Partenaires corporatifs**

Autorité des marchés financiers  
Banque de développement du Canada  
Banque du Canada  
Banque Laurentienne du Canada  
Banque Nationale du Canada  
Bell Canada  
BMO Groupe financier  
Caisse de dépôt et placement du Québec  
Fédération des caisses Desjardins du Québec  
Financière Sun Life, Québec  
Gaz Métro  
Hydro-Québec  
Industrie Canada  
Intact  
Investissements PSP  
Ministère des Finances et de l'Économie  
Power Corporation du Canada  
Rio Tinto Alcan  
Ville de Montréal

### **Partenaires universitaires**

École Polytechnique de Montréal  
École de technologie supérieure (ÉTS)  
HEC Montréal  
Institut national de la recherche scientifique (INRS)  
McGill University  
Université Concordia  
Université de Montréal  
Université de Sherbrooke  
Université du Québec  
Université du Québec à Montréal  
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

*This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.*

**ISSN 2292-0838 (en ligne)**

**Partenaire financier**

Enseignement supérieur,  
Recherche, Science  
et Technologie  
**Québec** 

# Horizon Effect in the Term Structure of Long-Run Risk-Return Trade-Offs

*Cédric Okou<sup>\*</sup>, Éric Jacquier<sup>†</sup>*

## Résumé/abstract

The horizon effect in the long-run predictive relationship between market excess return and historical market variance is investigated. To this end, the asymptotic multivariate distribution of the term structure of risk-return trade-offs is derived, accounting for short- and long-memory in the market variance dynamics. A rescaled Wald statistic is used to test whether the term structure of risk-return trade-offs is at, that is, the risk-return slope coefficients are equal across horizons. When the regression model includes an intercept, the premise of a at term structure of risk-return relationships is rejected. In contrast, there is no significant statistical evidence against the equality of slope coefficients from constrained risk-return regressions estimated at different horizons. A smoothed cross-horizon estimate is then proposed for the trade-off intensity at the market level. The findings underscore the importance of economically motivated restrictions to improve the estimation of intertemporal asset pricing models.

**Mots clés/keywords :** Horizon effect, Stock return predictability, Realized variance, Short-memory, Long-memory.

---

<sup>\*</sup> Corresponding address: Department of Finance, Université du Québec à Montréal, 315 rue Sainte-Catherine Est, Montréal, Québec, Canada. Tel.: +1 514 987 3000x5521; fax.: +1 514 987 0422.

E-mail address: okou.cedric@uqam.ca

<sup>†</sup> Department of Finance, HEC Montréal, Canada.

# 1. Introduction

For decades now, several studies have proposed models to forecast stock market returns based on empirical evidence and theoretical arguments. In practice, many investors track the time variation in asset returns to try to take advantage of predictability patterns and mispricings. Fama and French (1988) report early evidence of stock returns predictability induced by dividend yields over various time horizons. Since then, many variables including macroeconomic indicators, financial ratios and risk factors have widened the set of predictors that can explain a substantial part of the time variation in stock returns. See for example Merton (1980), Campbell and Shiller (1988), Boudoukh, Richardson and Whitelaw (2008).

In the return predictability framework, several models relate excess returns to risk measures. These models can be estimated at different horizons to yield a term structure of risk-return trade-offs. Variance measures emerge as good predictors of excess returns, delivering their full predictive power in the long-run as documented by Campbell and Viceira (2005), Bandi and Perron (hereafter BP) (2008), Bollerslev, Tauchen and Zhou (2009). At the market index level, a cursory look at the term structure of trade-offs between excess returns and historical variances, reveals a striking pattern. In general, the risk-return trade-off is hard to detect for short horizons but appears significant for longer horizons. The observed discrepancy between short versus long horizons trade-off estimates, illustrates the horizon effect. As pointed out by Camponovo, Scaillet and Trojani (2013), a few extreme observations might “hide” the evidence of short term return predictability. There is an ongoing debate on whether this empirical pattern is backed by an economic rationale or is just the result of a mere sampling artifact. Namely, the evidence of long horizon predictability in returns documented in Fama and French (1988), Valkanov (2003), Cochrane (2008), BP (2008) has been challenged by Stambaugh (1999), Ang and Bekaert (2007), Boudoukh et al. (2008), Goyal and Welch (2008) who raise the possibility of spurious long horizon relations.

The main purpose of this paper is to propose a formal detection test for the horizon effect in the predictive long-run relation between the market excess return and its historical variance. To fulfill this objective, we employ a methodology which explicitly accounts for persistence in the predictor, that is, the variance process. In their multi-horizon analysis of the risk-return trade-off, BP (2008) accommodate the persistence in the variance process using a near-unit-root specification. However, the variance dynamics can exhibit long-range memory as pointed out by Comte and Renault (1998), BP (2006), Corsi (2009), Bollerslev, Sizova

and Tauchen (2012), Sizova (2013) among others. Moreover, BP (2008) perform an individual treatment of the trade-off estimates while the inference on the term structure of predictive relationships should rely on joint test procedures to protect against spuriousness, as argued by Boudoukh et al. (2008). Because of the overlapping aggregation, there is a limited amount of independent information across horizons, and therefore, the estimated coefficients are strongly tied together. We propose a joint inference framework which is robust to the strong dependence among coefficients across horizons.

This article adds to the existing literature in two ways. First, we derive the joint asymptotic distribution of the term structure of risk-return trade-offs and the limiting approximation of its corresponding covariance matrix. We consider both local-to-unity (BP 2008) and long-memory (Sizova 2013) frameworks for the realized variance dynamics. Our results provide multivariate analytical tools to build appropriate Wald statistics and test some restrictions on the term structure of risk-return trade-offs. Thus, we generalize the analysis of BP (2008) and Sizova (2013) to the multivariate case and extend the results of Jacquier and Okou (2013) to long-memory predictors. Second, we exploit a standardized version of the classical Wald statistic, the *rescaled* Wald statistic, to assess the horizon effect. The proposed *rescaled* Wald statistic allows for multiple hypotheses testing while controlling for the level of type I error. In a multi-horizon context, this helps overcome the drawbacks of a “one-at-a-time” test strategy. We also provide simulated critical values for the *rescaled* Wald statistics.

Further, the inclusion of an intercept in the risk-return trade-off model plays a significant role. We find that the past market variance is positively related to the market excess return only at long horizons when an intercept is included in the model. By contrast, we cannot reject the assumption of no horizon effect for zero-intercept specifications as the estimated slope coefficients remain positive and have the same magnitude across all horizons. These results suggest a proportional risk-return trade-off relation consistent with Merton-type intertemporal capital asset pricing model (ICAPM). From the various horizon-specific estimates of the link between past variance and future excess return, we extract a smoothed proxy for the aggregate price of variance risk. We interpret this proxy as a “cross-horizon backward-looking” market price of risk. We obtain a reasonable value between 2 and 3.

The rest of this paper is structured as follows. We begin in Section 2 by introducing a simple risk-return model and discussing some estimation findings. Section 3 presents single inference results and simulated

critical values for the *rescaled*  $t$ -statistics. In Section 4, we perform a joint test based on the *rescaled* Wald statistic to investigate the horizon effect in the term structure of risk-return trade-offs. Specifically, we study the statistical equality between slope coefficients from different horizons. Thus, we characterize the term structure of risk-return trade-offs and relate it to Merton's (1973) ICAPM framework. Then, we draw on model averaging techniques to compute a cross-horizon estimate of the strength of the trade-off, a proxy for the market price of risk. We conclude in Section 5.

## 2. Framework

### 2.1 Model Specification

For the empirical work, we use NYSE/AMEX value-weighted index with dividends as the market proxy and 30-day T-bill rate as the risk-free rate. The data span the period from January 2, 1952 to December 31, 2009 and are retrieved from CRSP files. BP (2008) use the same dataset over a shorter period from January 2, 1952 through December 29, 2006. To begin our analysis, we specify the regressions

$$R_{t,t+h} = \alpha_h + \beta_h RV_{t-h,t} + \varepsilon_{t,t+h}, \quad (1)$$

where  $R_{t,t+h}$  is the market log-excess return between  $t$  and  $t+h$ ,  $RV_{t-h,t}$  denotes the past market variance between  $t-h$  and  $t$ , and  $\varepsilon_{t,t+h}$  represents a prediction error. The subscript  $h$  stands for the horizon and takes value between 1 and 120 months. The parameters  $\alpha_h$  and  $\beta_h$  are the intercept and the slope of the risk-return model. For a given horizon  $h$ ,  $\beta_h$  captures the trade-off intensity.

The baseline data are daily. Thus, we aggregate the daily continuously compounded return ( $r$ ) in excess of the risk-free rate ( $r^f$ ), to obtain the monthly log-excess return

$$R_{t,t+1} = \sum_{j=1}^{n_t} \left( r_{t+j/n_t} - r_{t+j/n_t}^f \right), \quad (2)$$

where  $n_t$  is the number of transaction days in month  $t$ . Monthly data are then used to construct  $h$ -month overlapping series. To compute the aggregated log-excess return series over  $(t, t+h)$ ,

$$R_{t,t+h} = \sum_{i=0}^{h-1} R_{t+i,t+1+i}, \quad (3)$$

we employ a  $h$ -length rolling window. An overlapping aggregation of the raw series is intended to produce longer and smoother aggregated series than a non-overlapping aggregation procedure. This common practice in long-run analysis is expected to generate more accurate estimates by reducing the noise. However, an overlapping aggregation might create some concern on the validity of traditional inference procedures when the resulting highly serially correlated variables are used within a least squares framework. See Campbell and Shiller (1988) or Boudoukh and Richardson (1993) for details.

We now discuss the computation of the regressor in the risk-return model of Eqn. (1). Realized variance measures are simple and consistent estimates for the second-order path variation of the returns based on infill asymptotics, i.e., asymptotics as the time distance between any two records shrinks. This justifies the use of higher frequency observations to compute quadratic variation estimates as popularized by Andersen, Bollerslev and Diebold (2010), Andersen, Bollerslev, Diebold and Labys (2003), Barndorff-Nielsen and Shephard (2003), among others. Specifically, the monthly variance  $RV_{t,t+1} = \sum_{j=1}^{n_t} r_{t+j/n_t}^2$  is estimated as the sum of squared daily log-returns where  $n_t$  is the number of daily observations available in month  $t$ . Similarly as the returns, for any horizon  $h$  longer than one month, we implement the aggregation according to

$$RV_{t,t+h} = \sum_{i=0}^{h-1} RV_{t+i,t+1+i}. \quad (4)$$

As the sampling frequency increases towards infinity ( $n_t \uparrow \infty$ ) Eqn. (4) yields a consistent estimator for the quadratic variation of the log-price process over  $(t, t+h)$ . Jacquier and Okou (2013) account for possible discontinuities in the return dynamics by separating continuous volatility from jumps in the risk-return trade-off analysis.

Other common variance measures such as realized kernels (Barndorff-Nielsen, Hansen, Lunde and Shephard 2008) may include Bartlett-type adjustments to account for autocorrelations in the daily returns. However, the effect of a potentially non-vanishing predictable component in the log-returns is relatively marginal in our sample.

Table 1 reports some descriptive statistics of realized log-excess returns ( $R_{t,t+1}$ ) and variances ( $RV_{t,t+1}$ ) for non-annualized monthly NYSE/AMEX value weighted market index from January 1952 to December 2009.

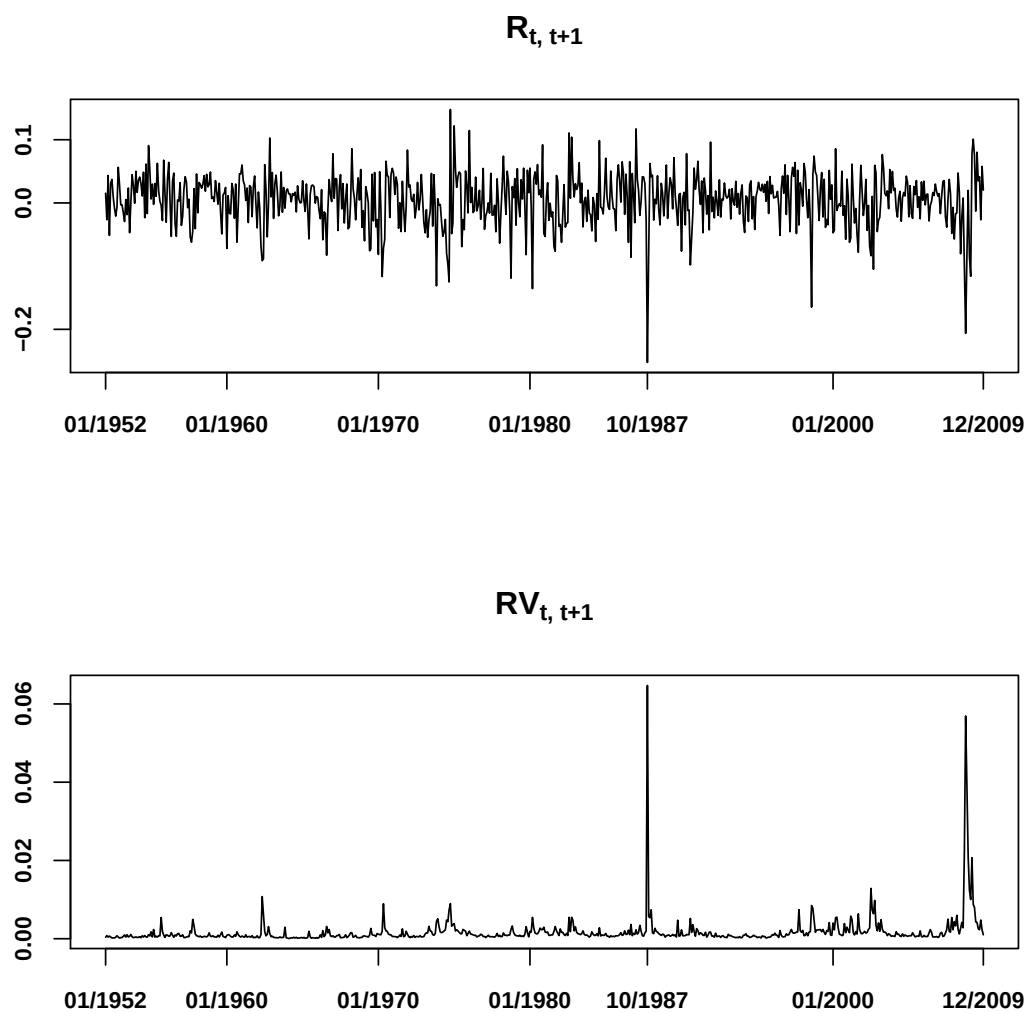
**Table 1**  
Descriptive statistics.

|                           | $R_{t,t+1}$ | $RV_{t,t+1}$ |
|---------------------------|-------------|--------------|
| Mean                      | 0.0044      | 0.0017       |
| Variance                  | 0.0018      | 1.644E-5     |
| Skewness                  | -0.803      | 10.778       |
| Kurtosis                  | 6.154       | 146.014      |
| Autocorrelation           | 0.098       | 0.503        |
| Autocorrelation 1952-1986 | 0.056       | 0.537        |
| Autocorrelation 1988-2007 | -0.006      | 0.502        |

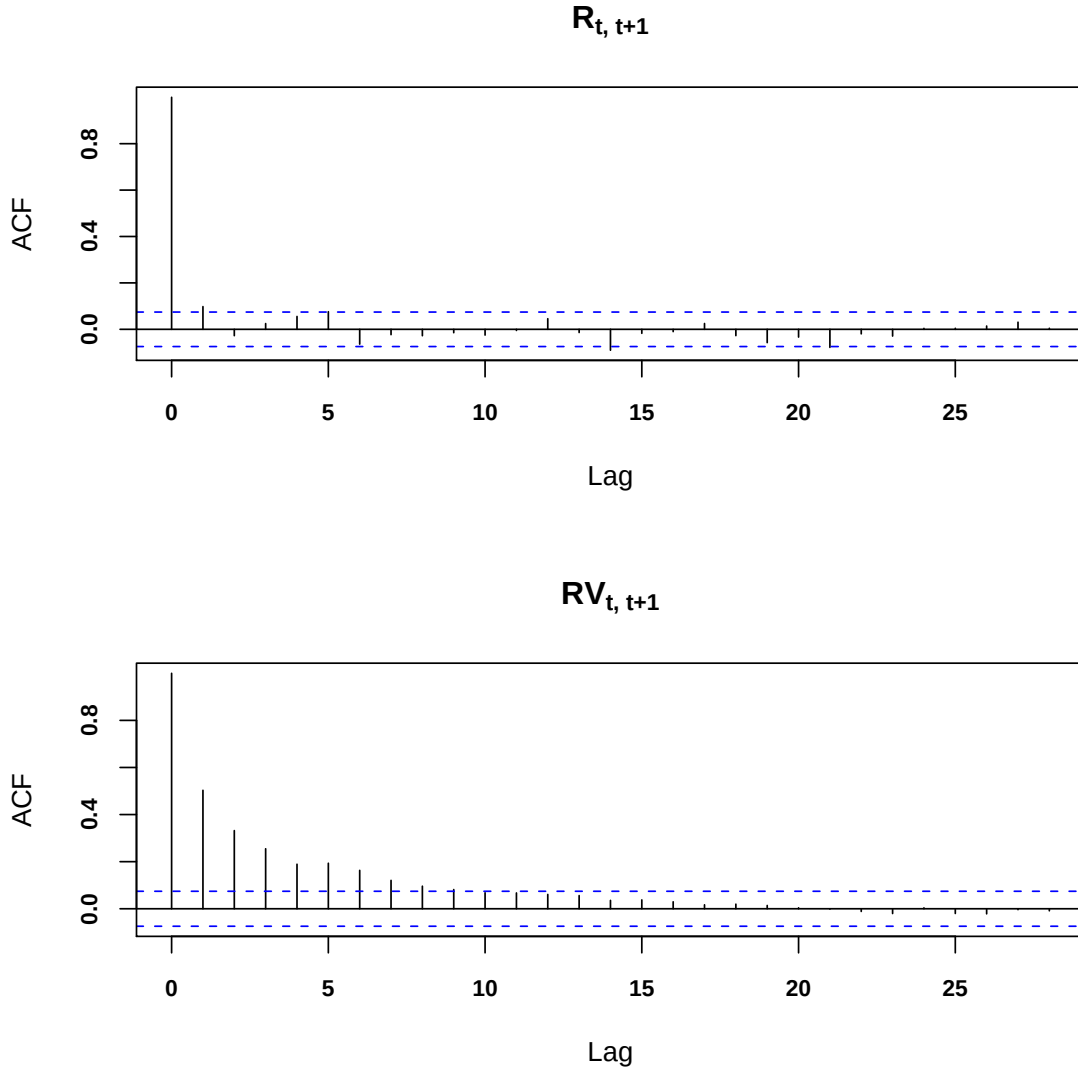
Typically, the market realized variance displays a right-skewed and fat-tailed distribution. This empirical regularity is discussed in Andersen, Bollerslev, Diebold and Labys (2003). Namely, the descriptive statistics show higher asymmetry and kurtosis for the realized variance series in contrast to the distribution of excess returns. A look at the dependence structure reveals a stronger first-lagged autocorrelation for  $RV_{t,t+1}$  (0.503) than for  $R_{t,t+1}$  (0.098) over the full sample. Table 1 also presents first-order autocorrelations in two subsamples excluding 1987 and post-2007 financial crisis periods. The dependence structure seems robust to the large swings in returns during crisis periods as subsample autocorrelation estimates are close to the full sample ones.

In Fig. 1, we plot the monthly market excess return ( $R_{t,t+1}$ ) and variance ( $RV_{t,t+1}$ ) time series. Clearly, the financial crash around October 1987 translates into an extreme quadratic variation activity. The end of the sample period also exhibits substantial variability related to the subprime crisis that started in 2007.





**Fig. 1.** Monthly NYSE / AMEX excess return (top) and realized variance (bottom), from Jan. 52 to Dec. 09.



**Fig. 2.** Autocorrelation of monthly excess return (top) and realized variance (bottom).

Fig. 2 presents the autocorrelation functions plots for  $R_{t,t+1}$  and  $RV_{t,t+1}$ . The bottom plot reveals some persistence in the distribution of the realized variance. Namely, we see significant autocorrelations for the realized variance series up to lag 6. The serial dependence in the quadratic variation activity is carried at least over a semester.

## 2.2 Evidence of Risk-Return Trade-Off

Predictability in returns is of prime importance for investment decisions and has become a centerpiece in empirical finance. Campbell (2001) points out that the persistence of the predictor and the strength of the short-run predictability are key determinants for long-run predictability. Moreover, the so-called *leverage effect* is expected to determine the intensity of the long-run risk-return relationship, see Stambaugh (1999). Interestingly, BP (2008) show that past market variance is a stronger predictor of future excess return than the dividend yield and the consumption-to-wealth ratio suggested by Lettau and Ludvigson (2001a). BP’s (2008) findings are robust in Valkanov’s (2003) local-to-unity framework.

We estimate the relationship between market excess returns and past market variances at different horizons ranging from 1 to 120 months. The regression results for Eqn. (1) are presented in the top panel of Table 2. The bottom panel reports the regression estimates when the intercept is constrained to zero. Note that in our empirical analysis, we use a sample which covers the 2007-2009 financial crisis period.

For now, we conduct inference according to the heteroskedasticity and autocorrelation consistent (HAC) correction of the standard *t-statistic*. It is important to mention that the HAC *t-statistic* is badly behaved, especially when the predictor is highly persistent due to overlapping aggregation. To address this issue, Hodrick (1992) and Nelson and Kim (1993) conduct Monte Carlo simulations supporting alternative inference procedures with improved small-sample properties. However, these simulation-based studies do not provide a systematic analytic answer within a unified inference framework, as pointed out by Valkanov (2003). Thus, contrasting or reconciling the conclusions from these simulations methods may be challenging, because their results highly depend on how the artificial data are being generated. Moreover, it might be difficult to pin down the potential effect of specific features of the data on the test statistics. In the next section, we propose a unified framework to conduct inference on the term structure of long horizon predictive regressions. Specifically, we derive the joint asymptotic distribution of the term structure of risk-return trade-offs and the limiting approximation of its corresponding covariance matrix, considering both local-to-unity (BP 2008) and long-memory (Sizova 2013) dynamics for the predictor. Our results yield multivariate analytical tools to construct adequate statistics and test some restrictions, including but not limited to predictability and horizon effect in the term structure of risk-return trade-offs.

Among our key findings, unconstrained regressions unveil statistically significant trade-offs only for ultra

**Table 2**

Trade-off estimates at different horizons for NYSE/AMEX 1952-2009. The regression  $R_{t,t+h} = \alpha_h + \beta_h RV_{t-h,t} + \varepsilon_{t,t+h}$  with and without intercept of  $h$ -period continuously compounded excess market returns  $R_{t,t+h} = \sum_{i=0}^{h-1} R_{t+i,t+1+i}$  on  $h$ -period past realized variances  $RV_{t-h,t} = \sum_{i=0}^{h-1} RV_{t-i-1,t-i}$  is run. Values of  $h = 1$  to 120 months are considered. Significance at the 5% level based on a two-sided test with HAC  $t$ -statistics is denoted by \*.

| h                                    | 1       | 3      | 6      | 12     | 24     | 36     | 48     | 60     | 72     | 84     | 96      | 108     | 120     |
|--------------------------------------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|---------|---------|
| With intercept                       |         |        |        |        |        |        |        |        |        |        |         |         |         |
| $\alpha_h$                           | 0.006   | 0.014  | 0.017  | 0.044  | 0.100  | 0.137  | 0.157  | 0.096  | -0.011 | -0.207 | -0.363  | -0.551  | -0.618  |
| HAC $t$                              | 4.005*  | 2.494* | 1.477  | 1.558  | 1.481  | 1.310  | 1.166  | 0.593  | -0.062 | -1.189 | -2.109* | -4.289* | -4.671* |
| $\beta_h$                            | -1.109  | -0.089 | 0.892  | 0.351  | -0.141 | 0.050  | 0.396  | 1.522  | 2.743  | 4.356  | 5.323   | 6.236   | 6.317   |
| HAC $t$                              | -2.350* | -0.145 | 2.669* | 0.473  | -0.113 | 0.037  | 0.324  | 1.200  | 1.890  | 3.545* | 5.006*  | 8.294*  | 8.697*  |
| $R^2$ (%)                            | 1.1     | 0.0    | 1.6    | 0.1    | 0.0    | 0.0    | 0.3    | 5.0    | 16.4   | 37.6   | 52.6    | 67.4    | 68.5    |
| Without intercept ( $\alpha_h = 0$ ) |         |        |        |        |        |        |        |        |        |        |         |         |         |
| $\beta_h$                            | -0.544  | 0.497  | 1.354  | 1.637  | 1.853  | 2.015  | 2.175  | 2.429  | 2.649  | 2.791  | 2.847   | 2.880   | 2.927   |
| HAC $t$                              | -0.839  | 0.673  | 2.999* | 2.841* | 3.52*  | 3.959* | 4.129* | 4.485* | 4.844* | 5.196* | 5.606*  | 5.803*  | 6.095*  |

long horizons (7 years and over).  $R^2$ s values range from 0.0% up to 68.5% and increase almost monotonically at longer horizons, consistently with BP's (2008) results.

The zero-intercept regressions show positive slopes which are statistically significant across all horizons, the only exceptions being monthly and quarterly horizons. Beyond the quarterly horizon, the risk-return trade-off estimates range from 1.4 to 3.0. It is important to notice that the zero-intercept constraint is an empirical refinement proposed in Campbell and Thompson (2008) for the ICAPM analysis. Imposing such a restriction yields economically meaningful trade-off intensities within a realistic range.

In a nutshell, the past market variance has a major contribution to the predictable part of excess returns at medium to long horizons. This is consistent with previous works suggesting historical volatility as a strong driver of return predictability at low frequencies. See Lettau and Ludvigson (2010) for a review.

In general, short term risk-return trade-offs are difficult to detect. Namely, short-run estimates are sensitive to short term variability of predictors. This is critical in the estimation of the risk-return relationship, especially when the data sample contains episodes of high volatility. For instance, monthly market variances during periods of financial instability in October 1987, Q4 2008 and Q1 2009 were too high, 40 to 50 times bigger than the trimmed average variance value for the whole post-war period. Exploring the relationship between expected stock returns and volatility, French, Schwert and Stambaugh (1987) discuss the challenge of finding a positive trade-off in the data at a short horizon. Volatility is known to be persistent and, as pointed out by Sizova (2013), performing inference with long-range dependent predictive variable is not trivial. A few additional data points can induce dramatic changes in the predictor's sample mean. Thus, the evidence of risk-return trade-off will be sensitive to shifts in the sample mean of volatility (see, e.g., Campbell and Yogo, 2006).

There are several attempts to improve the estimation of the risk-return trade-off. Ghysels, Santa-Clara and Valkanov (2005) suggest a mixed data sampling (or MIDAS) regression and conclude that "there is a risk-return trade-off after all". Maheu and McCurdy (2007) propose a truncated exponential realized volatility estimator, whereas DeGennaro and Zhao (1998) advocate for additional predictors in the risk-return model. Moreover, realized range-based variance estimators tend to outperform the sum of squared returns (realized variance) approximation of the integrated variance in terms of efficiency and robustness to market microstructure noise, as argued by Parkinson (1980), Beckers (1983), Alizadeh, Brandt and Diebold

(2002), among others. Christensen and Podolskij (2007) extend the homoscedastic diffusion assumption of Dijk and Martens (2007) to derive a general asymptotic theory supporting this evidence. Their theoretical results are confirmed by some empirical findings based on high-frequency data. However, the long-run risk-return analysis is often performed at low frequencies (typically 1 month horizon and beyond using not intra-day, but daily baseline data). Thus, the impact of microstructure contaminations on realized variance estimator in such a low frequency analysis should be minimal. This might explain why the long-run risk-return literature commonly adopt the realized variance as an estimate of the monthly (or longer horizon) variance.

An important line of criticism against the long-horizon predictability of returns is based upon the argument that standard inference may fail in the presence of a persistent predictor in the long-run regression. This failure of the classical asymptotics can manifest itself in terms of inconsistent estimators, test size distortions and convergence to functionals of Brownian motions, as discussed in Valkanov (2003). Moreover, Boudoukh et al. (2008) claim that short- and long-run estimators are almost perfectly correlated and therefore, econometricians should be “equally impressed by short- and long-horizon evidence” of returns predictability. We now provide an alternative asymptotic framework to address these issues.

### 3. Inferential Insights

In this section, we present test statistics with improved size properties in the presence of a persistent regressor. We assume both short- and long-memory frameworks for the realized variance dynamics. In our multi-horizon regressions, classical and HAC asymptotics lose their accuracy due to the overlapping aggregation. The observed size distortions, which increase with the horizons, are only partially corrected by HAC  $t$ -statistics. As an illustration, the bootstrap experiments conducted in BP (2008) show that the actual size is 3 times larger than the nominal size of 5% for a 10-year horizon. This leads to over-reading the evidence in favor of the alternative hypothesis of returns predictability in long-run estimations. Valkanov (2003) addresses this issue by proposing a standardized version of the classical  $t$ -statistic when the proportion of overlapping is a fixed fraction of the sample size. In this setting, the so-called *rescaled*  $t$ -statistic is a theoretically and empirically valid correction both for correlation and long horizon biases.

### 3.1 Local-to-Unity Asymptotics

To conduct inference on long-run risk-return trade-offs, BP (2008) use Valkanov's (2003) local-to-unity framework. Similarly, we consider the following system

$$R_{t,t+1} = \beta_1 RV_{t-1,t} + \varepsilon_{t,t+1}, \quad (5)$$

and

$$RV_{t,t+1} = \rho_0 + \rho_1 RV_{t-1,t} + u_{t,t+1}, \quad (6)$$

where Eqn. (5) describes the risk-return predictive relationship and Eqn. (6) sets a near-unit-root process for the predictor, the realized variance.

We summarize the assumptions for convenience.

**Assumptions** In model (5-6),

1.  $R_{t,t+h} = \sum_{i=0}^{h-1} R_{t+i,t+1+i}$  and  $RV_{t-h,t} = \sum_{i=0}^{h-1} RV_{t-i-1,t-i}$  are the long-horizon regressand and regressor.
2. The proportion of overlapping  $h = \lfloor \lambda T \rfloor$  is a non-trivial fraction of the sample size, where  $\lambda$  is fixed between 0 and 1, and  $\lfloor x \rfloor$  denotes the largest integer that is less than or equal to  $x$ .
3.  $(\varepsilon_{t,t+1}, u_{t,t+1})'$  is a vector martingale difference sequence with covariance matrix  $[\sigma_\varepsilon^2, \sigma_{\varepsilon u}, \bullet, \sigma_u^2]$  and finite fourth moments.
4.  $\rho_0 = 0$  and  $\rho_1 = 1 + c/T$ .

At least, two comments arise from the above assumptions. First, *Assumption 3* is more suitable when characterizing the evolution of the log-realized volatility. Second, deviations from the unit root for the autoregressive coefficient  $\rho_1$  in *Assumption 4*, are controlled by the parameter  $c$  at the decreasing rate  $T$ . Thus,  $\rho_1$  can be arbitrarily close to 1. Yet, the dynamics of the realized variance characterizes a stationary short-memory process.

To alleviate notation, we denote  $\lambda_{(a \vee b)} \equiv \max(\lambda_a, \lambda_b)$ . Further, we consider the following functionals of two processes  $X$  and  $Y$  that are defined and almost surely continuous over  $[0, 1]$ .

$$F_1(X_s, Y_s) \equiv \frac{\int_{\lambda}^{1-\lambda} X_s Y_s ds}{\int_{\lambda}^{1-\lambda} Y_s^2 ds},$$

$$F_2(X_s, Y_s) \equiv \frac{\int_{\lambda}^{1-\lambda} X_s^2 ds \int_{\lambda}^{1-\lambda} Y_s^2 ds - \left( \int_{\lambda}^{1-\lambda} X_s Y_s ds \right)^2}{\left( \int_{\lambda}^{1-\lambda} Y_s^2 ds \right)^2},$$

and

$$F_3(X_s(\lambda_a), X_s(\lambda_b), Y_s(\lambda_a), Y_s(\lambda_b)) \equiv$$

$$\frac{\int_{\lambda_{(a \vee b)}}^{1-\lambda_{(a \vee b)}} X_s(\lambda_a) X_s(\lambda_b) ds \int_{\lambda_{(a \vee b)}}^{1-\lambda_{(a \vee b)}} Y_s(\lambda_a) Y_s(\lambda_b) ds}{\int_{\lambda_a}^{1-\lambda_a} Y_s^2(\lambda_a) ds \int_{\lambda_b}^{1-\lambda_b} Y_s^2(\lambda_b) ds}$$

$$- \frac{\int_{\lambda_a}^{1-\lambda_a} X_s(\lambda_a) Y_s(\lambda_a) ds \int_{\lambda_b}^{1-\lambda_b} X_s(\lambda_b) Y_s(\lambda_b) ds \left( \int_{\lambda_{(a \vee b)}}^{1-\lambda_{(a \vee b)}} Y_s(\lambda_a) Y_s(\lambda_b) ds \right)^2}{\left( \int_{\lambda_a}^{1-\lambda_a} Y_s^2(\lambda_a) ds \right)^2 \left( \int_{\lambda_b}^{1-\lambda_b} Y_s^2(\lambda_b) ds \right)^2}.$$

We generalize the asymptotic results of BP (2008) in the propositions below. Namely, we extend BP's (2008) asymptotics to the multivariate case by considering the joint limiting distribution of slope coefficients estimated at two (w.l.o.g.) arbitrary horizons. It is straightforward to extend our results to more than two horizons.

**Proposition 1** (*The unrestricted regressions as per Eqn. (1)). For dynamics (5) and (6) under Assumptions (1-4),  $\beta_1 = 0$ , and Eqn. (1) estimated for any two arbitrary levels of aggregation  $h_i$  and  $h_j$  (w.l.o.g.),*

$$1. \ T \begin{pmatrix} \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor} \\ \hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} \end{pmatrix} \xrightarrow{\mathcal{L}} \frac{\sigma_{\varepsilon}}{\sigma_u} \begin{bmatrix} F_1(\overline{W}(s, \lambda_i), \overline{J}_c(s, -\lambda_i)) \\ F_1(\overline{W}(s, \lambda_j), \overline{J}_c(s, -\lambda_j)) \end{bmatrix},$$

$$2. \ T^3 Var \begin{pmatrix} \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor} \\ \hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} \overline{Avar}(\hat{\beta}_{h_i}) & \overline{Acov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) \\ \overline{Acov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) & \overline{Avar}(\hat{\beta}_{h_j}) \end{bmatrix},$$



where

$$\overline{Avar} \left( \widehat{\beta}_{h_l} \right) = \frac{\sigma_\varepsilon^2}{\sigma_u^2} \frac{1}{1-2\lambda_l} F_2 \left( \overline{W}(s, \lambda_l), \overline{J}_c(s, -\lambda_l) \right), \quad l = i, j,$$

$$\overline{Acov} \left( \widehat{\beta}_{h_i}, \widehat{\beta}_{h_j} \right) = \frac{\sigma_\varepsilon^2}{\sigma_u^2} \frac{1}{1-2\lambda_{(i \vee j)}} F_3 \left( \overline{W}(s, \lambda_i), \overline{W}(s, \lambda_j), \overline{J}_c(s, -\lambda_i), \overline{J}_c(s, -\lambda_j) \right),$$

$$\overline{W}(t, \lambda_l) = \{W(t + \lambda_l) - W(t)\} - \frac{1}{1-2\lambda_l} \int_{\lambda_l}^{1-\lambda_l} (W(s + \lambda_l) - W(s)) ds, \quad l = i, j,$$

$$\overline{J}_c(t, -\lambda_l) = \left\{ \int_{t-\lambda_l}^t J_c(s) ds \right\} - \frac{1}{1-2\lambda_l} \int_{\lambda_l}^{1-\lambda_l} \left( \int_{t-\lambda_l}^t J_c(s) ds \right) dt, \quad l = i, j,$$

$$dJ_c(s) = cJ_c(s) + dB(s), \quad J_c(0) = 0 \text{ or } J_c(s) = B(s) + c \int_0^s e^{c(s-\tau)} B(\tau) d\tau,$$

and  $\{W(s), B(s)\}$  is a vector of standard Brownian motions with covariance  $\sigma_{\varepsilon u} / \sigma_\varepsilon \sigma_u$ .

**Proposition 2** (The restricted regressions as per Eqn. (1) with  $\alpha_h = 0$ ). For dynamics (5) and (6) under Assumptions (1-4),  $\beta_1 = 0$ , and Eqn. (1) estimated with  $\alpha_h = 0$  for any two arbitrary levels of aggregation  $h_i$  and  $h_j$  (w.o.l.g),

$$1. \quad T \begin{pmatrix} \widehat{\beta}_{h_i = \lfloor \lambda_i T \rfloor} \\ \widehat{\beta}_{h_j = \lfloor \lambda_j T \rfloor} \end{pmatrix} \xrightarrow{\mathcal{L}} \frac{\sigma_\varepsilon}{\sigma_u} \begin{bmatrix} F_1(W(s, \lambda_i), J_c(s, -\lambda_i)) \\ F_1(W(s, \lambda_j), J_c(s, -\lambda_j)) \end{bmatrix},$$

$$2. \quad T^3 Var \begin{pmatrix} \widehat{\beta}_{h_i = \lfloor \lambda_i T \rfloor} \\ \widehat{\beta}_{h_j = \lfloor \lambda_j T \rfloor} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} Avar(\widehat{\beta}_{h_i}) & Acov(\widehat{\beta}_{h_i}, \widehat{\beta}_{h_j}) \\ Acov(\widehat{\beta}_{h_i}, \widehat{\beta}_{h_j}) & Avar(\widehat{\beta}_{h_j}) \end{bmatrix},$$

where

$$Avar \left( \widehat{\beta}_{h_l} \right) = \frac{\sigma_\varepsilon^2}{\sigma_u^2} \frac{1}{1-2\lambda_l} F_2 \left( W(s, \lambda_l), J_c(s, -\lambda_l) \right), \quad l = i, j,$$

$$Acov \left( \widehat{\beta}_{h_i}, \widehat{\beta}_{h_j} \right) = \frac{\sigma_\varepsilon^2}{\sigma_u^2} \frac{1}{1-2\lambda_{(i \vee j)}} F_3 \left( W(s, \lambda_i), W(s, \lambda_j), J_c(s, -\lambda_i), J_c(s, -\lambda_j) \right),$$

$$W(t, \lambda_l) = W(t + \lambda_l) - W(t), \quad l = i, j,$$

$$J_c(t, -\lambda_l) = \int_{t-\lambda_l}^t J_c(s) ds, \quad l = i, j,$$

$$dJ_c(s) = cJ_c(s) + dB(s), \quad J_c(0) = 0 \quad \text{or} \quad J_c(s) = B(s) + c \int_0^s e^{c(s-\tau)} B(\tau) d\tau,$$

and  $\{W(s), B(s)\}$  is a vector of standard Brownian motions with covariance  $\sigma_{\varepsilon u} / \sigma_{\varepsilon} \sigma_u$ .

The above propositions provide key ingredients for constructing adequate test statistics and inspecting their asymptotic properties. The *rescaled*  $t$ -statistic, used to test for the significance of the slope coefficient in Eqn. (1), has a well-defined asymptotic distribution. See Propositions 1 and 2 in BP (2008). That is, for the unrestricted regressions,

$$\frac{t_{\hat{\beta}_{h=\lfloor \lambda T \rfloor}}}{\sqrt{T}} \xrightarrow{\mathcal{L}} \frac{F_1(\overline{W}(s, \lambda), \overline{J}_c(s, -\lambda))}{\sqrt{\frac{1}{1-2\lambda} F_2(\overline{W}(s, \lambda), \overline{J}_c(s, -\lambda))}},$$

and for the restricted regressions  $\alpha_h = 0$ ,

$$\frac{t_{\hat{\beta}_{h=\lfloor \lambda T \rfloor}}}{\sqrt{T}} \xrightarrow{\mathcal{L}} \frac{F_1(W(s, \lambda), J_c(s, -\lambda))}{\sqrt{\frac{1}{1-2\lambda} F_2(W(s, \lambda), J_c(s, -\lambda))}}.$$

### 3.2 Long-Memory Asymptotics

While the autocorrelation of the returns is insignificant beyond lag 0, the return variance series exhibits persistence which lasts over many periods. Thus, the realized variance may reveal a long-memory pattern reflecting in a slow decay of its correlogram. Recall that for a long-range dependent process, the first autocorrelations are not necessarily high but taper off at a hyperbolic rate. BP (2006) find evidence of long-memory in the realized and implied volatility processes and argue for a fractional cointegrating relation between the two series. Sizova (2013) studies the small-sample effects arising in long-horizon regressions, assuming short- and long-memory variance dynamics. Now, consider the following system

$$R_{t,t+1} = \beta_1 Z_t(d) + \vartheta_t, \tag{7}$$

and

$$RV_{t,t+1} = Z_{t+1}(d), \quad (8)$$

where  $\vartheta_t$  is a short-memory process and  $Z_t(d)$  denotes a long-memory process.

A basic long-memory model for  $Z_{t+1}(d)$  is the fractionally integrated Gaussian noise. It is defined as the unique stationary solution of the difference equations  $(1 - L)^d Z_{t+1}(d) = u_{t,t+1}$ , with  $u_{t,t+1} \sim N(0, \sigma_u^2)$  and the fractional parameter  $d \in ]-1/2, 1/2[$ . The condition  $d > -1/2$  ensures invertibility while  $d < 1/2$  guarantees stationarity. Fractionally integrated models raise many challenges. Namely, the fractional difference filter can introduce some artificial mixing between long and short term characteristics, as reported by Comte and Renault (1998). Moreover, the implementation of this class of process necessitates a long build-up period.

In general, the definition of a long-memory process either characterizes the decay rates of its long-lag autocorrelations or the explosion rates of its low-frequency spectral density in the neighborhood of zero. The conditions for the equivalence between these two definitions can be found in Beran (1994). Based on the autocorrelation structure, a covariance stationary process  $Z_t(d)$  has long-memory if the condition  $\lim_{h \rightarrow \infty} \text{corr}(Z_{t+h}(d), Z_t(d)) = Gh^{2d-1}$  with  $G \neq 0$ , holds. This definition seems more intuitive as pointed out by Breidt, Crato and de Lima (1998).

Alternatively, a long-memory process  $Z_t(d)$  has an unbounded spectral density  $f_z(\delta)$  in a neighborhood of  $\delta = 0$ , and there exists a positive constant  $\kappa$  such that  $\lim_{\delta \rightarrow 0} f_z(\delta) \|1 - e^{-i\delta}\|^{2d} = \kappa$ , for some  $0 < d < 1/2$ . In addition,  $Z_t(d)$  has a finite variance and therefore,  $\int_{-\pi}^{\pi} f_z(\delta) d\delta < \infty$ . This definition focuses on the stationary subclass of long-memory processes. It excludes non-stationary processes which are obtained when  $d \geq 1/2$ . When the differencing parameter  $d$  controlling the order of integration is equal to 0, the spectral density defines a short-memory process. A detailed discussion is presented in Tsay and Chung (2000).

In some cases, the clear identification of long-memory can be delicate. Diebold and Inoue (2001) argue that nonzero but “small” amounts of structural change may be mistaken for long-memory.

The previous characterizations of a long-memory process are more general than the one we consider based on the asymptotic distribution of its partial sums (Baillie, 1996). We rely on the following assumptions.

**Assumptions** In model (7-8),

5.  $Z_t(d)$  is a long-memory process of order  $d \in ]0, 1/2[$  and  $\vartheta_t$  is a short-memory process.

6. For a given  $\lambda \in [0, 1]$  and a non-negative differencing parameter  $d$ ,

$$\frac{\sum_{s=1}^{h=\lfloor \lambda T \rfloor} [Z_s(d) - E(Z_s(d))]}{T^{1/2+d}} \xrightarrow{\mathcal{L}} \sigma_d B_d(\lambda),$$

where  $E$  is the expectation operator,  $\lim_{T \rightarrow \infty} \frac{\text{Var}(\sum_{s=1}^T Z_s(d))}{T^{1+2d}} = \sigma_d^2$ ,

$$B_d(t) \equiv \sqrt{\frac{(1+2d)\Gamma(1-d)}{\Gamma(1+d)\Gamma(1-2d)}} \left\{ \begin{array}{l} \int_0^t (t-s)^d dB(s) \\ + \int_{-\infty}^0 [(t-s)^d - (-s)^d] dB(s) \end{array} \right\}$$

is a type I fractional Brownian motion, and  $B(t)$  is a standard Brownian motion.

Clearly, when  $d = 0$ ,  $B_0(t)$  is identical to the standard Brownian motion  $B(t)$ . Type I fractional Brownian motion has emerged in the paper by Mandelbrot and Van Ness (1968) and has become the cornerstone of many subsequent works such as Taqqu (1975), Gripenberg and Norros (1996), Tsay and Chung (2000), Marinucci and Robinson (2000), Davidson and Hashimzade (2009), and others.

As for the local-to-unity case, we derive the joint asymptotic distribution of slope coefficients estimated at two (w.l.o.g.) different horizons when the predictor exhibits long-memory dependence. Thus, as presented in the propositions below, we provide a multivariate extension of the asymptotic results in Sizova (2013).

**Proposition 3** (The unrestricted regressions as per Eqn. (1)). For dynamics (7) and (8) under Assumptions (1-2, 5-6),  $\beta_1 = 0$ , and Eqn. (1) estimated for any two arbitrary levels of aggregation  $h_i$  and  $h_j$  (w.l.o.g.),

$$1. \ T^d \begin{pmatrix} \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor} \\ \hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} \end{pmatrix} \xrightarrow{\mathcal{L}} \frac{\sigma_{\vartheta}}{\sigma_d} \begin{bmatrix} F_1(\overline{W}(s, \lambda_i), \overline{B}_d(s, -\lambda_i)) \\ F_1(\overline{W}(s, \lambda_j), \overline{B}_d(s, -\lambda_j)) \end{bmatrix},$$

$$2. \ T^{1+2d} \text{Var} \begin{pmatrix} \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor} \\ \hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} \overline{Avar}(\hat{\beta}_{h_i}) & \overline{Acov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) \\ \overline{Acov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) & \overline{Avar}(\hat{\beta}_{h_j}) \end{bmatrix},$$

where

$$\overline{Avar}(\hat{\beta}_{h_l}) = \frac{\sigma_{\vartheta}^2}{\sigma_d^2} \frac{1}{1-2\lambda_l} F_2(\overline{W}(s, \lambda_l), \overline{B}_d(s, -\lambda_l)), \quad l = i, j,$$

$$\overline{Acov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) = \frac{\sigma_{\vartheta}^2}{\sigma_d^2} \frac{1}{1-2\lambda_{(i \vee j)}} F_3(\overline{W}(s, \lambda_i), \overline{W}(s, \lambda_j), \overline{B}_d(s, -\lambda_i), \overline{B}_d(s, -\lambda_j)),$$

$$\overline{W}(t, \lambda_l) = \{W(t + \lambda_l) - W(t)\} - \frac{1}{1-2\lambda_l} \int_{\lambda_l}^{1-\lambda_l} (W(s + \lambda_l) - W(s)) ds, \quad l = i, j,$$

$$\overline{B}_d(s, -\lambda_l) = \{B_d(t) - B_d(t - \lambda_l)\} - \frac{1}{1-2\lambda_l} \int_{\lambda_l}^{1-\lambda_l} (B_d(s) - B_d(s - \lambda_l)) ds, \quad l = i, j,$$

$$B_d(s) = \sqrt{\frac{(1+2d)\Gamma(1-d)}{\Gamma(1+d)\Gamma(1-2d)}} \left\{ \int_0^s (s - \tau)^d dB(\tau) + \int_{-\infty}^0 [(s - \tau)^d - (-\tau)^d] dB(\tau) \right\},$$

and  $\{W(s), B(s)\}$  is a vector of standard Brownian motions with covariance  $\sigma_{\vartheta d}/\sigma_{\vartheta}\sigma_d$ .

**Proposition 4** (The restricted regressions as per Eqn. (1) with  $\alpha_h = 0$ ). For dynamics (7) and (8) under Assumptions (1-2, 5-6),  $\beta_1 = 0$ , and Eqn. (1) estimated with  $\alpha_h = 0$  for any two arbitrary levels of aggregation  $h_i$  and  $h_j$  (w.l.o.g.),

$$1. \ T^d \begin{pmatrix} \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor} \\ \hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} \end{pmatrix} \xrightarrow{\mathcal{L}} \frac{\sigma_{\vartheta}}{\sigma_d} \begin{bmatrix} F_1(W(s, \lambda_i), B_d(s, -\lambda_i)) \\ F_1(W(s, \lambda_j), B_d(s, -\lambda_j)) \end{bmatrix},$$

$$2. \ T^{1+2d} \text{Var} \begin{pmatrix} \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor} \\ \hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{bmatrix} Avar(\hat{\beta}_{h_i}) & Acov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) \\ Acov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) & Avar(\hat{\beta}_{h_j}) \end{bmatrix},$$

where

$$Avar(\hat{\beta}_{h_l}) = \frac{\sigma_{\vartheta}^2}{\sigma_d^2} \frac{1}{1-2\lambda_l} F_2(W(s, \lambda_l), B_d(s, -\lambda_l)), \quad l = i, j,$$

$$Acov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) = \frac{\sigma_{\vartheta}^2}{\sigma_d^2} \frac{1}{1-2\lambda_{(i \vee j)}} F_3(W(s, \lambda_i), W(s, \lambda_j), B_d(s, -\lambda_i), B_d(s, -\lambda_j)),$$

$$W(t, \lambda_l) = W(t + \lambda_l) - W(t), \quad l = i, j,$$

$$B_d(s, -\lambda_l) = B_d(t) - B_d(t - \lambda_l), \quad l = i, j,$$

$$B_d(s) = \sqrt{\frac{(1+2d)\Gamma(1-d)}{\Gamma(1+d)\Gamma(1-2d)}} \left\{ \int_0^s (s - \tau)^d dB(\tau) + \int_{-\infty}^0 \left[ (s - \tau)^d - (-\tau)^d \right] dB(\tau) \right\},$$

and  $\{W(s), B(s)\}$  is a vector of standard Brownian motions with covariance  $\sigma_{\vartheta d}/\sigma_{\vartheta}\sigma_d$ .

In the long-memory framework of system (7)-(8), we can also derive the asymptotic distribution of the *rescaled*  $t$ -statistic used to test the significance of the slope coefficients estimated in Eqn. (1).

For the unrestricted regressions,

$$\frac{t_{\hat{\beta}_{h=\lfloor \lambda T \rfloor}}}{\sqrt{T}} \xrightarrow{\mathcal{L}} \frac{F_1(\overline{W}(s, \lambda), \overline{B}_d(s, -\lambda))}{\sqrt{\frac{1}{1-2\lambda} F_2(\overline{W}(s, \lambda), \overline{B}_d(s, -\lambda))}},$$

and for the restricted regressions  $\alpha_h = 0$ ,

$$\frac{t_{\hat{\beta}_{h=\lfloor \lambda T \rfloor}}}{\sqrt{T}} \xrightarrow{\mathcal{L}} \frac{F_1(W(s, \lambda), B_d(s, -\lambda))}{\sqrt{\frac{1}{1-2\lambda} F_2(W(s, \lambda), B_d(s, -\lambda))}}.$$

Finally, in the context of overlapping aggregation of predicted and predictive series, the *rescaled* Wald statistic is appropriate to overcome the issues arising with the joint inference for long-run regressions. The *rescaled* Wald statistic has very appealing sample properties while usual Wald statistics tend to diverge. In addition, the *rescaled* Wald statistic provides a proper answer to the adjustments for multiple hypotheses testing.

For a given restriction  $\mathcal{G}$ , the *rescaled* Wald test statistic

$$T^{-1} \chi_{(\mathcal{G})}^2 = T^{-1} \left\{ \left( \mathcal{G} \hat{\underline{\beta}} \right)' \left[ \mathcal{G} Var \left( \hat{\underline{\beta}} \right) \mathcal{G}' \right]^{-1} \left( \mathcal{G} \hat{\underline{\beta}} \right) \right\}, \quad (9)$$

converges to a non-degenerate distribution. Note that the vector  $\hat{\underline{\beta}}' = (\hat{\beta}_{h_1}, \dots, \hat{\beta}_{h_K})$  is the estimator of the term structure of risk-return trade-offs. The matrix  $Var(\hat{\underline{\beta}})$  denotes the corresponding covariance matrix of the  $\hat{\underline{\beta}}$  estimator with typical elements  $Var(\hat{\beta}_{h_i}, \hat{\beta}_{h_j})$  defined in the Appendix. The *rescaled* Wald

statistic can be used to investigate the joint significance of the set of slope coefficients. To test for the null hypothesis of no predictability ( $\beta_{h_1} = \beta_{h_2} = \dots = \beta_{h_K} = 0$ ), one can use the identity matrix  $\mathbf{I}$  to set the restriction. Boudoukh et al. (2008) use a standard Wald statistic to test for joint significance, since they do not consider a high degree of persistence such as near-unit-root or long-memory dynamics for the regressor.

Eqn. (9) also provides a metric to assess the relative importance of the horizon effect in multiple horizons risk-return regressions. One can test if the risk-return relationship is proportional across horizons, that is, if the slope coefficients are all equal.

### 3.3 Simulation Results

We simulate the asymptotic distribution of the *rescaled*  $t$ -statistic to obtain its critical values. The simulations are implemented under the null of no predictability, with a first set of parameter values chosen to mimic the observed data. We consider both short- and long-memory dynamics for the realized variance process. The simulation parameters are  $R_t = \varepsilon_t$ ,  $RV_t = \rho_1 RV_{t-1} + u_t$ ,  $\text{corr}(\varepsilon_t, u_t) = -0.3$ ,  $\sigma_\varepsilon = 1$ ,  $\sigma_u = 1$  for the short-memory (SM) case and  $R_t = \varepsilon_t$ ,  $RV_t = (1 - L)^{-d} u_t$ ,  $\text{corr}(\varepsilon_t, u_t) = -0.3$ ,  $\sigma_\varepsilon = 1$ ,  $\sigma_u = 1$  for the long-memory (LM) case. The simulation results for the autoregressive coefficient  $\rho_1$  and the fractional parameter  $d$  matching their empirical estimates of 0.6 and 0.44 are displayed in Table 3a. For robustness check, we also simulate the critical values for different designs of the memory parameters ( $\rho_1 = 0.3, 0.9$  and  $d = 0.2$ ) corresponding to different levels of persistence in the realized variance dynamics, as shown in Table 3b.

For the slope estimates previously reported in Table 2, we calculate their corresponding *rescaled*  $t$ -statistics  $t_{\hat{\beta}_h}/\sqrt{T}$ . As these empirical *rescaled*  $t$ -statistics increase with the horizon, their simulated distributions under the null hypothesis become more disperse, yielding wider confidence intervals. This pattern is observed whether we specify a short- or long-memory dynamics for the predictive variable  $RV$ . A wild bootstrap procedure implemented in BP (2008) yields similar critical values. For the unconstrained regressions, the *rescaled*  $t$ -statistics fall outside the simulated confidence bounds beyond 7 years. This suggests a long-run predictability as previously found with the classical HAC inference. By contrast, the zero-intercept regressions reveal significant trade-offs beyond the quarterly frequency. Interestingly, long-range dependence

**Table 3a**

Empirical estimates and simulated critical values for *rescaled*  $t$ -statistics under no predictability. Short-memory and long-memory dependence for the market realized variance are considered. Simulation parameters are  $R_t = \varepsilon_t$ ,  $RV_t = 0.6RV_{t-1} + u_t$ ,  $\text{corr}(\varepsilon_t, u_t) = -0.3$ ,  $\sigma_\varepsilon = 1$ ,  $\sigma_u = 1$  for the short-memory (SM) case and  $R_t = \varepsilon_t$ ,  $RV_t = (1 - L)^{-0.44} u_t$ ,  $\text{corr}(\varepsilon_t, u_t) = -0.3$ ,  $\sigma_\varepsilon = 1$ ,  $\sigma_u = 1$  for the long-memory (LM) case. Continuously compounded market excess returns and market realized variances are simulated under the assumption of no predictability. 10,000 replications of 696 monthly observations are constructed. Then, the simulated series are aggregated over  $h$  periods (1 to 120 months) and  $h$ -horizon predictive regressions (with and without an intercept) of continuously compounded market excess returns on past market variances are run. Empirical estimates of  $t_{\hat{\beta}_h}/\sqrt{T}$  and critical values simulated from the asymptotic distributions are reported. For instance, “SM 97.5% c.v.” is the right-tail simulated critical value of  $t_{\hat{\beta}_h}/\sqrt{T}$  for a 5% two-sided test with equal probability of rejection in the tails given a short-memory specification for the realized variance. [\*] indicates significant *rescaled*  $t$ -statistic empirical values outside SM 97.5% c.v. simulated bounds.

| h                                    | 1      | 3      | 6        | 12       | 24       | 36       | 48       | 60       | 72       | 84       | 96       | 108      | 120      |
|--------------------------------------|--------|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| With intercept                       |        |        |          |          |          |          |          |          |          |          |          |          |          |
| $t_{\hat{\beta}_h}/\sqrt{T}$         | -0.106 | -0.011 | 0.127    | 0.036    | -0.015   | 0.006    | 0.054    | 0.208    | 0.394    | 0.675    | 0.896[*] | 1.193[*] | 1.193[*] |
| LM 97.5% c.v.                        | 0.077  | 0.131  | 0.183    | 0.261    | 0.380    | 0.472    | 0.552    | 0.630    | 0.704    | 0.770    | 0.856    | 0.920    | 0.995    |
| SM 97.5% c.v.                        | 0.075  | 0.122  | 0.168    | 0.239    | 0.340    | 0.414    | 0.485    | 0.555    | 0.627    | 0.697    | 0.787    | 0.855    | 0.925    |
| SM 2.5% c.v.                         | -0.074 | -0.115 | -0.162   | -0.215   | -0.297   | -0.369   | -0.427   | -0.463   | -0.516   | -0.561   | -0.605   | -0.639   | -0.682   |
| LM 2.5% c.v.                         | -0.073 | -0.121 | -0.166   | -0.239   | -0.345   | -0.423   | -0.512   | -0.584   | -0.646   | -0.724   | -0.793   | -0.859   | -0.924   |
| Without intercept ( $\alpha_h = 0$ ) |        |        |          |          |          |          |          |          |          |          |          |          |          |
| $t_{\hat{\beta}_h}/\sqrt{T}$         | -0.056 | 0.069  | 0.226[*] | 0.243[*] | 0.348[*] | 0.500[*] | 0.636[*] | 0.787[*] | 0.989[*] | 1.200[*] | 1.318[*] | 1.379[*] | 1.433[*] |
| LM 97.5% c.v.                        | 0.075  | 0.125  | 0.176    | 0.251    | 0.361    | 0.450    | 0.525    | 0.607    | 0.671    | 0.734    | 0.805    | 0.877    | 0.952    |
| SM 97.5% c.v.                        | 0.073  | 0.119  | 0.163    | 0.230    | 0.323    | 0.393    | 0.461    | 0.517    | 0.570    | 0.636    | 0.690    | 0.738    | 0.783    |
| SM 2.5% c.v.                         | -0.074 | -0.117 | -0.164   | -0.221   | -0.309   | -0.386   | -0.448   | -0.499   | -0.550   | -0.604   | -0.660   | -0.701   | -0.771   |
| LM 2.5% c.v.                         | -0.074 | -0.123 | -0.173   | -0.254   | -0.356   | -0.438   | -0.510   | -0.584   | -0.659   | -0.736   | -0.800   | -0.870   | -0.930   |



**Table 3b**

Simulated critical values of *rescaled*  $t$ -statistics for different designs of memory parameters. Short-memory and long-memory dependence for the market realized variance are considered. Simulation parameters are  $R_t = \varepsilon_t$ ,  $RV_t = \rho_1 RV_{t-1} + u_t$ ,  $\text{corr}(\varepsilon_t, u_t) = -0.3$ ,  $\sigma_\varepsilon = 1$ ,  $\sigma_u = 1$  for the short-memory (SM) case and  $R_t = \varepsilon_t$ ,  $RV_t = (1 - L)^{-d} u_t$ ,  $\text{corr}(\varepsilon_t, u_t) = -0.3$ ,  $\sigma_\varepsilon = 1$ ,  $\sigma_u = 1$  for the long-memory (LM) case. Continuously compounded market excess returns and market realized variances are simulated under the assumption of no predictability. 10,000 replications of 696 monthly observations are constructed. Then, the simulated series are aggregated over  $h$  periods (1 to 120 months) and  $h$ -horizon predictive regressions (with and without an intercept) of continuously compounded market excess returns on past market variances are run. Critical values simulated from the asymptotic distributions are reported.

| h                                    | 1      | 3      | 6      | 12     | 24     | 36     | 48     | 60     | 72     | 84     | 96     | 108    | 120    |
|--------------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| With intercept                       |        |        |        |        |        |        |        |        |        |        |        |        |        |
| SM $\rho_1 = 0.3$                    |        |        |        |        |        |        |        |        |        |        |        |        |        |
| 97.5% c.v.                           | 0.076  | 0.116  | 0.161  | 0.226  | 0.322  | 0.397  | 0.476  | 0.540  | 0.609  | 0.676  | 0.749  | 0.819  | 0.882  |
| 2.5% c.v.                            | -0.074 | -0.113 | -0.157 | -0.215 | -0.295 | -0.360 | -0.414 | -0.472 | -0.523 | -0.569 | -0.613 | -0.656 | -0.693 |
| SM $\rho_1 = 0.9$                    |        |        |        |        |        |        |        |        |        |        |        |        |        |
| 97.5% c.v.                           | 0.077  | 0.131  | 0.182  | 0.257  | 0.359  | 0.436  | 0.528  | 0.595  | 0.664  | 0.715  | 0.781  | 0.859  | 0.941  |
| 2.5% c.v.                            | -0.071 | -0.122 | -0.166 | -0.236 | -0.320 | -0.397 | -0.459 | -0.516 | -0.556 | -0.599 | -0.650 | -0.691 | -0.733 |
| LM $d = 0.2$                         |        |        |        |        |        |        |        |        |        |        |        |        |        |
| 97.5% c.v.                           | 0.075  | 0.116  | 0.167  | 0.243  | 0.357  | 0.446  | 0.527  | 0.619  | 0.704  | 0.779  | 0.847  | 0.914  | 0.990  |
| 2.5% c.v.                            | -0.072 | -0.112 | -0.155 | -0.214 | -0.305 | -0.377 | -0.436 | -0.502 | -0.545 | -0.591 | -0.654 | -0.696 | -0.741 |
| Without intercept ( $\alpha_h = 0$ ) |        |        |        |        |        |        |        |        |        |        |        |        |        |
| SM $\rho_1 = 0.3$                    |        |        |        |        |        |        |        |        |        |        |        |        |        |
| 97.5% c.v.                           | 0.075  | 0.115  | 0.159  | 0.221  | 0.311  | 0.375  | 0.444  | 0.501  | 0.552  | 0.613  | 0.665  | 0.726  | 0.775  |
| 2.5% c.v.                            | -0.075 | -0.114 | -0.157 | -0.220 | -0.300 | -0.377 | -0.436 | -0.495 | -0.542 | -0.589 | -0.644 | -0.689 | -0.745 |
| SM $\rho_1 = 0.9$                    |        |        |        |        |        |        |        |        |        |        |        |        |        |
| 97.5% c.v.                           | 0.075  | 0.127  | 0.178  | 0.248  | 0.341  | 0.414  | 0.475  | 0.542  | 0.596  | 0.638  | 0.703  | 0.759  | 0.818  |
| 2.5% c.v.                            | -0.072 | -0.125 | -0.172 | -0.242 | -0.329 | -0.409 | -0.474 | -0.523 | -0.585 | -0.634 | -0.689 | -0.740 | -0.777 |
| LM $d = 0.2$                         |        |        |        |        |        |        |        |        |        |        |        |        |        |
| 97.5% c.v.                           | 0.074  | 0.112  | 0.161  | 0.236  | 0.332  | 0.419  | 0.493  | 0.572  | 0.639  | 0.691  | 0.747  | 0.808  | 0.872  |
| 2.5% c.v.                            | -0.074 | -0.115 | -0.162 | -0.228 | -0.327 | -0.407 | -0.472 | -0.540 | -0.592 | -0.657 | -0.730 | -0.783 | -0.847 |

in  $RV$  tends to favor the null hypothesis of no predictability in returns. That is, a short-memory dynamics in  $RV$  produces tighter confidence intervals as compared to a long-memory specification. This confirms the findings of BP (2008) and Sizova (2013).

With these observations in mind, we now turn to the joint test of no predictability  $H_0 : \beta_{h_1} = \beta_{h_2} = \dots = \beta_{h_K} = 0$  versus  $H_1 : \beta_{h_i} \neq 0$  for some horizon  $h_i$ . The calculated *rescaled* Wald statistic  $T^{-1}\chi^2_{(I)} = 38.357$  for the unconstrained regressions (6.621 for the constrained regressions) is greater than its simulated 5% right-tail critical value of 2.303 (1.404). These results provide some global evidence against the null hypothesis of no predictability in returns.

## 4. Investigating the Horizon Effect

This section investigates the evidence of the horizon effect in the term structure of risk-return trade-offs. Recall that the term structure of risk-return trade-offs is the set of slope coefficients  $(\beta_{h_1}, \dots, \beta_{h_K})$  for the entire spectrum of horizons. From our previous findings, realized variance explains a substantial part of medium to long term market excess returns. For the unconstrained regressions, we observe that past market variances are positively related to market excess returns only for long horizons over 7 years. However, the term structure of risk-return trade-offs appears “nearly” flat for zero-intercept specifications as the estimated slope coefficients remain positive and have similar magnitudes at all horizons beyond 3 months. Following Lanne and Saikkonen (2006), BP (2008) argue that the zero-intercept constraint is a useful restriction yielding slope estimates with increased accuracy. Thus, the inclusion of an intercept term can compromise the efficient analysis of the horizon effect in the term structure of risk-return trade-offs.

### 4.1 Testing for Restrictions on the Term Structure

The origin of the term structure of risk-return trade-offs can be found in the dynamic changes to the investment opportunities. Campbell and Viceria (2005) introduce a simple model of return dynamics to compute the term structure of risk-return trade-offs from U.S. post-war quarterly data. Here, we want to assess whether the slope coefficients are similar from an econometric standpoint. That is, if the slopes are not statistically distinct from one another, the resulting term structure of risk-return trade-offs is immune to

horizon effect. Otherwise, there is a substantial horizon effect in the linkage between market excess return and past market realized variance.

### Pairwise Comparison of Slope Coefficients: one-at-a-time strategy

In the light of the previous asymptotical derivations, we construct statistics for testing some continuous restrictions on a vector of slopes fitted at different horizons. Namely, we compute *rescaled* Wald statistics for testing significant differences between slope coefficients from different horizons. The *rescaled* Wald statistic provides a useful way to explore the presence of sizable spreads between the slopes.

We now introduce two four-argument functionals  $\mathcal{N}(\cdot)$  and  $\mathcal{D}(\cdot)$ , as the building blocks of the asymptotic distribution of the *rescaled* Wald test statistic. These functionals are rigorously defined in the Appendix. We invoke the Functional Central Limit Theorem (FCLT) and the Continuous Mapping Theorem (CMT) to derive our asymptotics.

**Corollary 1** (*Local-to-unity asymptotics*).

1. For the unrestricted regressions,

$$T^{-1}\chi^2_{\left(\hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} - \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor}\right)} \xrightarrow{\mathcal{L}} \frac{\mathcal{N}\left(\overline{W}(s, \lambda_i), \overline{W}(s, \lambda_j), \overline{J}_c(s, -\lambda_i), \overline{J}_c(s, -\lambda_j)\right)}{\mathcal{D}\left(\overline{W}(s, \lambda_i), \overline{W}(s, \lambda_j), \overline{J}_c(s, -\lambda_i), \overline{J}_c(s, -\lambda_j)\right)}.$$

2. For the restricted regressions  $\alpha_h = 0$ ,

$$T^{-1}\chi^2_{\left(\hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} - \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor}\right)} \xrightarrow{\mathcal{L}} \frac{\mathcal{N}\left(W(s, \lambda_i), W(s, \lambda_j), J_c(s, -\lambda_i), J_c(s, -\lambda_j)\right)}{\mathcal{D}\left(W(s, \lambda_i), W(s, \lambda_j), J_c(s, -\lambda_i), J_c(s, -\lambda_j)\right)}.$$

**Corollary 2** (*Long-memory asymptotics*).

1. For the unrestricted regressions,

$$T^{-1}\chi^2_{\left(\hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} - \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor}\right)} \xrightarrow{\mathcal{L}} \frac{\mathcal{N}(\overline{W}(s, \lambda_i), \overline{W}(s, \lambda_j), \overline{B}_d(s, -\lambda_i), \overline{B}_d(s, -\lambda_j))}{\mathcal{D}(\overline{W}(s, \lambda_i), \overline{W}(s, \lambda_j), \overline{B}_d(s, -\lambda_i), \overline{B}_d(s, -\lambda_j))}.$$

2. For the restricted regressions  $\alpha_h = 0$ ,

$$T^{-1}\chi^2_{\left(\hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} - \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor}\right)} \xrightarrow{\mathcal{L}} \frac{\mathcal{N}(W(s, \lambda_i), W(s, \lambda_j), B_d(s, -\lambda_i), B_d(s, -\lambda_j))}{\mathcal{D}(W(s, \lambda_i), W(s, \lambda_j), B_d(s, -\lambda_i), B_d(s, -\lambda_j))}.$$

Interestingly, the asymptotic distribution of the *rescaled* Wald statistic for testing the difference between a pair of slopes from two distinct horizons, converges to a non-degenerate distribution. It is worth mentioning that test size distortions might arise from a multiple pairwise comparisons strategy. To assess the horizon effect, performing multiple pairwise equality tests often requires a Bonferroni-type of correction to control for the level of type I error. One can easily generalize the pairwise test results to more than two horizons.

### Joint Test: global strategy

To try to fully answer the question of whether the term structure of risk-retrun trade-offs is immune to horizon effect, we assess the statistical proximity inside the set of slope coefficients estimated for the whole horizon spectrum. To this end, we test the equality of slopes across horizons  $H_0 : \beta_{h_1} = \beta_{h_2} = \dots = \beta_{h_K}$  versus  $H_1 : \beta_{h_i} \neq \beta_{h_j}$ , for some horizons  $h_i$  and  $h_j$ . Lack of evidence against  $\beta_{h_1} = \beta_{h_2} = \dots = \beta_{h_K}$  will point to a flat term structure of risk-retrun trade-offs. Our strategy consists in constructing a test statistic using the sum of squared differences  $\hat{\beta}_{h_i} - \hat{\beta}_{h_1}$  for  $i = 2, \dots, K$ , adequately standardized. We rely on the *rescaled* Wald test statistic  $T^{-1}\chi^2_{(\Delta)}$  where the restriction matrix is

$${}_{(K-1) \times K} \Delta = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -1 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

## 4.2 Critical Values

In this subsection, we use daily observations to illustrate the asymptotic results for the detection of the horizon effect. Therefore, we do not refer to asymptotics literally. Nevertheless, we expect that running medium to long term analysis from daily data is a good practice for understanding medium- to long-run financial market dynamics.

To support our theoretical results, we report the *rescaled* Wald statistics for testing pairwise equality between slope coefficients from NYSE/AMEX regressions as per Eqn. (1) in the panel (a) of Table 4. The panel (b) in Table 4 provides the simulated critical values for  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  at 5% level of probability. To obtain the critical values, we generate 10,000 paths of 696 monthly observations of continuously compounded excess returns and realized variances, consistently with the actual sample size. The simulations are performed under the assumption of no predictability. Simulation parameters are  $R_t = \varepsilon_t$ ,  $RV_t = 0.6RV_{t-1} + u_t$ ,  $\text{corr}(\varepsilon_t, u_t) = -0.3$ ,  $\sigma_\varepsilon = 1$ ,  $\sigma_u = 1$  for the short-memory (SM) specification and  $R_t = \varepsilon_t$ ,  $RV_t = (1 - L)^{-0.44}u_t$ ,  $\text{corr}(\varepsilon_t, u_t) = -0.3$ ,  $\sigma_\varepsilon = 1$ ,  $\sigma_u = 1$  for the long-memory (LM) specification. These parameters are chosen to replicate the empirical properties of the actual data as much as possible. The autoregressive coefficient of the realized variance dynamics is consistent with the parameter value used by BP (2008) in their local-to-unity analysis. Moreover, BP (2006) report long-memory parameter estimates of the realized volatility process for the S&P 100 market index between 0.443 and 0.473 using a local Whittle estimator. Finally, the negative correlation between return and variance innovations is a well-documented stock market feature, sometimes referred to as the *leverage effect* as discussed in Jacquier, Polson and Rossi (2004). The simulated series are then aggregated over  $h$  periods (1 to 120 months) and  $h$ -horizon predictive regressions with and without an intercept, of excess returns on past realized variances are estimated. Simulated critical values of  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  for a one-sided test of size 5% using Bonferroni adjustment are reported. Significant spreads between  $\hat{\beta}_{h_i}$  and  $\hat{\beta}_{h_j}$  should generate  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  test statistic values in the panel (a) of Table 4 that are larger than their corresponding critical values in the panel (b) of Table 4. Note that we only present short-memory simulations since long-range variance dynamics yield similar results for pairwise equality tests.

When the regressions include an intercept, the observed differences are not statistically significant at

short/medium horizon, i.e., up to 7 years. The estimates of the pairwise equality test statistics  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  for  $h_i = 1, \dots, 72$  and  $h_j = 3, \dots, 84$  in the panel (a) of Table 4, are all included in their simulated 95% confidence intervals in the panel (b) of Table 4. This reflects the fact that slope coefficients are found statistically insignificant, thus mutually close for horizons between 1 to 84 months. Consistently with the detected long-run risk-return relationship, we see significant  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  values in the bottom left rectangular area of the panel (a) in Table 4, for  $h_i = 1, \dots, 84$  and  $h_j = 96, 108, 120$ . Finally, even though slopes become significant beyond 7 years, their values stay close from an econometric standpoint. Therefore, the pairwise equality test statistics are found insignificant in the bottom right corner of the panel (a) in Table 4, for  $h_i = 96, 108, 120$  and  $h_j = 108, 120$ .

**Table 4**

Pairwise equality test statistics  $T^{-1}\chi^2_{(\hat{\beta}_{h_j} - \hat{\beta}_{h_i})}$  for Eqn. (1).

Panel a:  $h$ -horizon predictive regressions, with an intercept, of continuously compounded NYSE/AMEX value-weighted excess returns from 1952 to 2009 on past realized variances as per Eqn. (1), are run. Horizons  $h = 1$  to 120 months are considered and  $T^{-1}\chi^2_{(\hat{\beta}_{h_j} - \hat{\beta}_{h_i})}$  estimates are reported. [\*] indicates a significant *rescaled* Wald statistic value outside the simulated confidence bounds shown in Panel b.

Panel b: simulation parameters are  $R_t = \varepsilon_t$ ,  $RV_t = 0.6RV_{t-1} + u_t$ ,  $corr(\varepsilon_t, u_t) = -0.3$ ,  $\sigma_\varepsilon = 1$ ,  $\sigma_u = 1$ . 10,000 samples of 696 months of continuously compounded excess returns and realized variances are simulated under the assumption of no predictability. Then, the simulated series are aggregated over  $h$  periods (1 to 120 months) and  $h$ -horizon predictive regressions with an intercept, of excess returns on past realized variances are run. Simulated critical values of  $T^{-1}\chi^2_{(\hat{\beta}_{h_j} - \hat{\beta}_{h_i})}$  for a one-sided test of size 5% using Bonferroni correction are reported.

| Panel a: Estimates of the pairwise equality test statistics                 |          |          |          |          |          |          |          |          |          |          |       |       |
|-----------------------------------------------------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-------|-------|
| $h_j=1$                                                                     | $h_i=1$  | 3        | 6        | 12       | 24       | 36       | 48       | 60       | 72       | 84       | 96    | 108   |
| 3                                                                           | 0.004    |          |          |          |          |          |          |          |          |          |       |       |
| 6                                                                           | 0.009    | 0.008    |          |          |          |          |          |          |          |          |       |       |
| 12                                                                          | 0.010    | 0.001    | 0.002    |          |          |          |          |          |          |          |       |       |
| 24                                                                          | 0.005    | 0.000    | 0.008    | 0.001    |          |          |          |          |          |          |       |       |
| 36                                                                          | 0.008    | 0.000    | 0.006    | 0.000    | 0.000    |          |          |          |          |          |       |       |
| 48                                                                          | 0.014    | 0.002    | 0.002    | 0.000    | 0.002    | 0.001    |          |          |          |          |       |       |
| 60                                                                          | 0.043    | 0.022    | 0.004    | 0.010    | 0.020    | 0.020    | 0.013    |          |          |          |       |       |
| 72                                                                          | 0.075    | 0.071    | 0.035    | 0.040    | 0.064    | 0.067    | 0.057    | 0.016    |          |          |       |       |
| 84                                                                          | 0.177    | 0.155    | 0.133    | 0.119    | 0.157    | 0.176    | 0.152    | 0.090    | 0.031    |          |       |       |
| 96                                                                          | 0.289[*] | 0.294[*] | 0.234[*] | 0.239[*] | 0.244[*] | 0.284[*] | 0.273[*] | 0.170[*] | 0.093[*] | 0.033[*] |       |       |
| 108                                                                         | 0.399[*] | 0.436[*] | 0.369[*] | 0.286[*] | 0.359[*] | 0.420[*] | 0.412[*] | 0.279[*] | 0.184[*] | 0.082[*] | 0.014 |       |
| 120                                                                         | 0.405[*] | 0.441[*] | 0.379[*] | 0.310[*] | 0.364[*] | 0.426[*] | 0.417[*] | 0.273[*] | 0.218[*] | 0.155[*] | 0.016 | 0.000 |
| Panel b: Simulated critical values of the pairwise equality test statistics |          |          |          |          |          |          |          |          |          |          |       |       |
| $h_j=1$                                                                     | $h_i=1$  | 3        | 6        | 12       | 24       | 36       | 48       | 60       | 72       | 84       | 96    | 108   |
| 3                                                                           | 0.006    |          |          |          |          |          |          |          |          |          |       |       |
| 6                                                                           | 0.011    | 0.009    |          |          |          |          |          |          |          |          |       |       |
| 12                                                                          | 0.015    | 0.021    | 0.017    |          |          |          |          |          |          |          |       |       |
| 24                                                                          | 0.026    | 0.043    | 0.050    | 0.036    |          |          |          |          |          |          |       |       |
| 36                                                                          | 0.038    | 0.065    | 0.078    | 0.075    | 0.044    |          |          |          |          |          |       |       |
| 48                                                                          | 0.050    | 0.084    | 0.102    | 0.108    | 0.077    | 0.046    |          |          |          |          |       |       |
| 60                                                                          | 0.063    | 0.105    | 0.129    | 0.141    | 0.126    | 0.090    | 0.044    |          |          |          |       |       |
| 72                                                                          | 0.076    | 0.124    | 0.153    | 0.173    | 0.170    | 0.128    | 0.095    | 0.045    |          |          |       |       |
| 84                                                                          | 0.183    | 0.168    | 0.177    | 0.205    | 0.211    | 0.185    | 0.166    | 0.099    | 0.045    |          |       |       |
| 96                                                                          | 0.104    | 0.168    | 0.206    | 0.232    | 0.241    | 0.221    | 0.182    | 0.155    | 0.083    | 0.028    |       |       |
| 108                                                                         | 0.120    | 0.193    | 0.237    | 0.264    | 0.283    | 0.272    | 0.250    | 0.212    | 0.165    | 0.071    | 0.042 |       |
| 120                                                                         | 0.138    | 0.219    | 0.270    | 0.304    | 0.322    | 0.317    | 0.301    | 0.256    | 0.201    | 0.123    | 0.098 | 0.041 |

**Table 5**

Pairwise equality test statistics  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  for Eqn. (1) estimated without an intercept.

Panel a:  $h$ -horizon predictive regressions, without an intercept, of continuously compounded NYSE/AMEX value-weighted excess returns from 1952 to 2009, on past realized variances as per Eqn. (1) are run. Horizons  $h = 1$  to 120 months are considered and  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  estimates are reported. [\*] indicates a significant *rescaled* Wald statistic value outside the simulated confidence bounds shown in Panel b.

Panel b: simulation parameters are  $R_t = \varepsilon_t$ ,  $RV_t = 0.6RV_{t-1} + u_t$ ,  $\text{corr}(\varepsilon_t, u_t) = -0.3$ . 10,000 samples of 696 months of continuously compounded excess returns and realized variances are simulated under the assumption of no predictability. Then, the simulated series are aggregated over  $h$  periods (1 to 120 months) and  $h$ -horizon predictive regressions without an intercept, of excess returns on past realized variances are run. Simulated critical values of  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  for a one-sided test of size 5% using Bonferroni correction are reported.

| Panel a: Estimates of the pairwise equality test statistics |          |          |       |       |       |       |       |       |       |       |       |       |
|-------------------------------------------------------------|----------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                                                             | $h_i=1$  | 3        | 6     | 12    | 24    | 36    | 48    | 60    | 72    | 84    | 96    | 108   |
| $h_j=1$                                                     |          |          |       |       |       |       |       |       |       |       |       |       |
| 3                                                           | 0.006    |          |       |       |       |       |       |       |       |       |       |       |
| 6                                                           | 0.028[*] | 0.009[*] |       |       |       |       |       |       |       |       |       |       |
| 12                                                          | 0.034[*] | 0.019[*] | 0.001 |       |       |       |       |       |       |       |       |       |
| 24                                                          | 0.048[*] | 0.034[*] | 0.004 | 0.001 |       |       |       |       |       |       |       |       |
| 36                                                          | 0.060[*] | 0.041[*] | 0.009 | 0.003 | 0.001 |       |       |       |       |       |       |       |
| 48                                                          | 0.072[*] | 0.056[*] | 0.015 | 0.005 | 0.003 | 0.002 |       |       |       |       |       |       |
| 60                                                          | 0.087[*] | 0.071[*] | 0.026 | 0.012 | 0.010 | 0.008 | 0.006 |       |       |       |       |       |
| 72                                                          | 0.102[*] | 0.086[*] | 0.039 | 0.019 | 0.020 | 0.018 | 0.017 | 0.005 |       |       |       |       |
| 84                                                          | 0.113[*] | 0.104[*] | 0.050 | 0.026 | 0.024 | 0.030 | 0.027 | 0.013 | 0.003 |       |       |       |
| 96                                                          | 0.119[*] | 0.121[*] | 0.055 | 0.030 | 0.028 | 0.031 | 0.027 | 0.016 | 0.005 | 0.001 |       |       |
| 108                                                         | 0.138[*] | 0.135[*] | 0.056 | 0.030 | 0.030 | 0.032 | 0.027 | 0.015 | 0.005 | 0.001 | 0.000 |       |
| 120                                                         | 0.145[*] | 0.149[*] | 0.060 | 0.032 | 0.033 | 0.035 | 0.031 | 0.015 | 0.007 | 0.002 | 0.001 | 0.000 |

| Panel b: Simulated critical values of the pairwise equality test statistics |         |       |       |       |       |       |       |       |       |       |       |       |
|-----------------------------------------------------------------------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                                                                             | $h_i=1$ | 3     | 6     | 12    | 24    | 36    | 48    | 60    | 72    | 84    | 96    | 108   |
| $h_j=1$                                                                     |         |       |       |       |       |       |       |       |       |       |       |       |
| 3                                                                           | 0.007   |       |       |       |       |       |       |       |       |       |       |       |
| 6                                                                           | 0.008   | 0.007 |       |       |       |       |       |       |       |       |       |       |
| 12                                                                          | 0.015   | 0.017 | 0.016 |       |       |       |       |       |       |       |       |       |
| 24                                                                          | 0.025   | 0.031 | 0.048 | 0.034 |       |       |       |       |       |       |       |       |
| 36                                                                          | 0.036   | 0.038 | 0.073 | 0.070 | 0.041 |       |       |       |       |       |       |       |
| 48                                                                          | 0.047   | 0.052 | 0.094 | 0.100 | 0.072 | 0.041 |       |       |       |       |       |       |
| 60                                                                          | 0.058   | 0.067 | 0.117 | 0.130 | 0.112 | 0.080 | 0.040 |       |       |       |       |       |
| 72                                                                          | 0.070   | 0.078 | 0.139 | 0.157 | 0.153 | 0.112 | 0.081 | 0.040 |       |       |       |       |
| 84                                                                          | 0.084   | 0.097 | 0.162 | 0.184 | 0.185 | 0.153 | 0.120 | 0.083 | 0.039 |       |       |       |
| 96                                                                          | 0.095   | 0.113 | 0.183 | 0.208 | 0.215 | 0.195 | 0.151 | 0.129 | 0.084 | 0.038 |       |       |
| 108                                                                         | 0.110   | 0.124 | 0.215 | 0.240 | 0.246 | 0.232 | 0.200 | 0.166 | 0.129 | 0.082 | 0.036 |       |
| 120                                                                         | 0.125   | 0.144 | 0.243 | 0.270 | 0.276 | 0.268 | 0.241 | 0.201 | 0.174 | 0.129 | 0.080 | 0.034 |

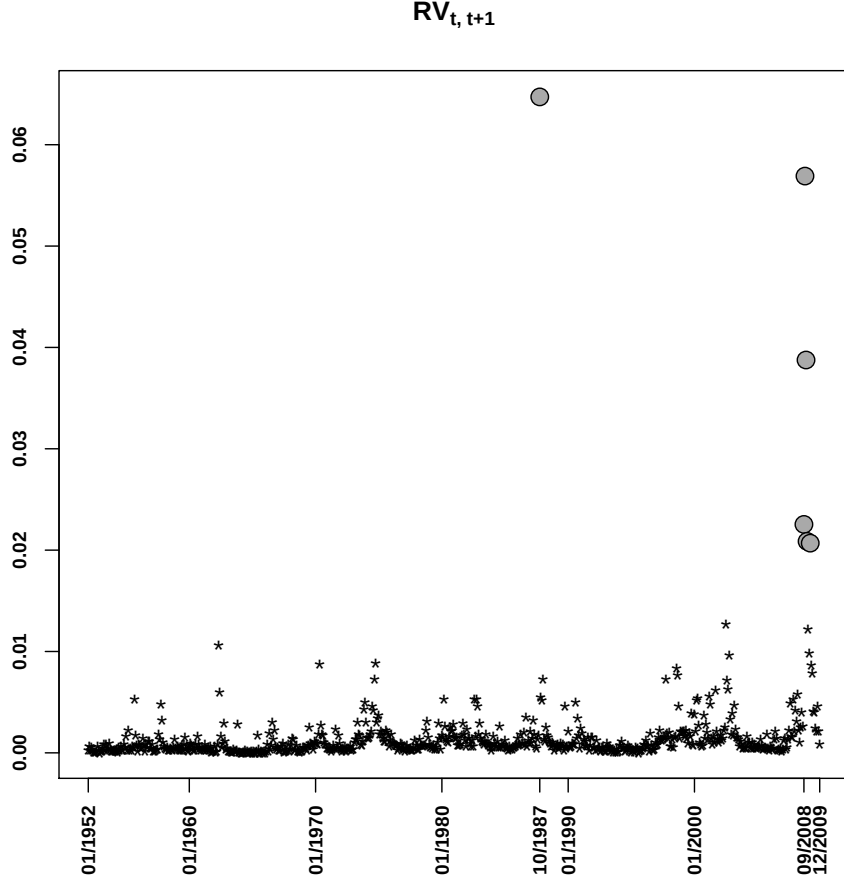


Constraining the intercept to zero in the regressions yields significant, positive and statistically close slopes across horizons beyond the quarterly frequency. As a consequence, the panel (a) in Table 5 exhibits significant pairwise equality test values only when  $h_i = 1, 3$  and  $h_j = 6, \dots, 120$ . For constrained regressions, we fail to obtain evidence against a flat risk-return term structure beyond 3 months.

### 4.3 Sensitivity of short term estimates to the variability of the predictor

Looking at the significant difference between monthly and quarterly constrained regression estimates, and the values obtained at longer horizons, one could conclude that there is some evidence of horizon effect. We consider otherwise since the stability and robustness of short term estimations are overwhelmingly affected by extreme volatility records, especially with the two major financial crises in our sampling window. Camponovo, Scaillet and Trojani (2013) use a similar argument to propose a robust bootstrap method for testing predictive relations that might be distorted by few anomalous observations. As shown in Fig. 3, monthly variances for October 1987, Q4 2008 and Q1 2009 appear extremely high. These extreme values are suspected to drive the negative and/or insignificant trade-offs found at the short term horizon.

To assess the sensitivity of short-run estimates to the variability of the predictor, we windsorize the realized variance series. The primary benefit of pre-processing the variance series is to generate more robust and stable estimations across horizons. The windsorization does not trim some observations away from the sample, but rather decreases the magnitude of the extreme variance values by substituting them with their 99% quantile. Obviously, data windsorization may not be adequate for some risk monitoring purposes. When monthly variance data are windsorized ( $\widetilde{RV}$ ) before the aggregation, the sluggish short term (monthly and quarterly) trade-off intensities present sizeable increases to become positive and statistically significant for constrained intercept regressions.



**Fig. 3.** Monthly realized variance with 1% largest values plotted as dark grey filled circles.

Table 6 reports a striking increase in the monthly (respectively quarterly) slope value from -0.544 (respectively 0.497) to 1.775 (respectively 2.409) for zero-intercept regressions. A percentage point rise in the monthly (quarterly) realized variance induces an average increase in the excess return on the next month (3 months) of 1.8% (2.4%). These numbers are quite comparable to the 2.9% increase at 10-years horizon. Thus, sampling variation may explain the weak short term trade-offs in the data. Beyond the quarterly horizon, the distorting effect of extreme variance values becomes less important in the regressions.

As shown in Table 7, the difference between slopes becomes statistically insignificant across all horizons including the monthly and quarterly horizons. Accordingly, the *rescaled* Wald test statistic for a joint restriction test  $T^{-1}\chi^2_{(\Delta)} = 0.061$  is found smaller than its simulated 5% right-tail critical value of 0.724. This reflects a lack of evidence against the assumption of a flat term structure of risk-return trade-offs.

**Table 6**

Trade-off estimates at different horizons for zero-intercept regressions with 99% windorized variance series for NYSE/AMEX 1952-2009. The regression  $R_{t,t+h} = \alpha_h + \beta_h \widetilde{RV}_{t-h,t} + \varepsilon_{t,t+h}$  with and without an intercept of  $h$ -period continuously compounded market excess returns  $\widetilde{R}_{t,t+h} = \sum_{i=0}^{h-1} R_{t+i,t+1+i}$  on  $h$ -period past windorized realized variances  $\widetilde{RV}_{t-h,t} = \sum_{i=0}^{h-1} \widetilde{RV}_{t-i-1,t-i}$  is run. Monthly 99% windorized variance series  $\widetilde{RV}_{t-1,t}$  are aggregated at horizons  $h = 1$  to 120 months. Significance at the 5% level based on a two-sided test with HAC  $t$ -statistics is denoted by \*. Significance at the 5% level based on a two-sided test with *rescaled*  $t$ -statistics  $t/\sqrt{T}$  is denoted by [\*].

| h                                    | 1        | 3        | 6        | 12       | 24       | 36       | 48       | 60       | 72       | 84       | 96       | 108      | 120      |
|--------------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| With intercept                       |          |          |          |          |          |          |          |          |          |          |          |          |          |
| $\alpha_h$                           | 0.004    | 0.008    | 0.023    | 0.060    | 0.103    | 0.137    | 0.158    | 0.095    | -0.037   | -0.220   | -0.374   | -0.562   | -0.620   |
| HAC $t$                              | 1.820    | 1.029    | 1.477    | 1.508    | 1.305    | 1.289    | 1.163    | 0.582    | -0.199   | -1.226   | -2.139*  | -4.334*  | -4.700*  |
| $\beta_h$                            | 0.149    | 1.228    | 0.404    | -0.614   | -0.231   | 0.035    | 0.387    | 1.531    | 3.016    | 4.485    | 5.418    | 6.331    | 6.337    |
| HAC $t$                              | 0.081    | 0.682    | 0.231    | -0.348   | -0.143   | 0.025    | 0.313    | 1.169    | 2.087*   | 3.499*   | 4.934*   | 8.142*   | 8.692*   |
| $t/\sqrt{T}$                         | 0.005    | 0.053    | 0.021    | -0.040   | -0.022   | 0.004    | 0.052    | 0.205    | 0.425    | 0.687    | 0.903[*] | 1.197[*] | 1.194[*] |
| $R^2$ (%)                            | 0.0      | 0.3      | 0.0      | 0.2      | 0.1      | 0.0      | 0.3      | 4.8      | 18.5     | 38.4     | 53.0     | 67.6     | 68.5     |
| Without intercept ( $\alpha_h = 0$ ) |          |          |          |          |          |          |          |          |          |          |          |          |          |
| $\beta_h$                            | 1.775    | 2.409    | 2.196    | 1.939    | 2.023    | 2.039    | 2.182    | 2.445    | 2.702    | 2.808    | 2.857    | 2.887    | 2.928    |
| HAC $t$                              | 1.340    | 1.914    | 1.902    | 2.357*   | 3.443*   | 3.877*   | 4.096*   | 4.519*   | 5.033*   | 5.156*   | 5.529*   | 5.727*   | 6.077*   |
| $t/\sqrt{T}$                         | 0.079[*] | 0.164[*] | 0.192[*] | 0.228[*] | 0.358[*] | 0.502[*] | 0.636[*] | 0.786[*] | 1.010[*] | 1.203[*] | 1.316[*] | 1.373[*] | 1.431[*] |

**Table 7**

Estimates of pairwise equality test statistics  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  for zero-intercept regressions with 99% windsorized variance series.  $h$ -horizon predictive regressions, without an intercept, of continuously compounded NYSE/AMEX value-weighted excess returns from 1952 to 2009 on 99% windsorized realized variance series as per Eqn. (1), are run. Horizons  $h = 1$  to 120 months are considered and  $T^{-1}\chi^2_{(\hat{\beta}_{h_j}-\hat{\beta}_{h_i})}$  estimates are reported. [\*] indicates a significant *rescaled* Wald statistic value outside the simulated confidence bounds shown in the Panel b in Table 5.

|         | $h_i=1$ | 3     | 6     | 12    | 24    | 36    | 48    | 60    | 72    | 84    | 96    | 108   |
|---------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $h_j=1$ |         |       |       |       |       |       |       |       |       |       |       |       |
| 3       | 0.001   |       |       |       |       |       |       |       |       |       |       |       |
| 6       | 0.000   | 0.000 |       |       |       |       |       |       |       |       |       |       |
| 12      | 0.000   | 0.001 | 0.000 |       |       |       |       |       |       |       |       |       |
| 24      | 0.000   | 0.001 | 0.000 | 0.000 |       |       |       |       |       |       |       |       |
| 36      | 0.000   | 0.001 | 0.000 | 0.000 | 0.000 |       |       |       |       |       |       |       |
| 48      | 0.000   | 0.000 | 0.000 | 0.001 | 0.001 | 0.001 |       |       |       |       |       |       |
| 60      | 0.001   | 0.000 | 0.000 | 0.003 | 0.005 | 0.008 | 0.006 |       |       |       |       |       |
| 72      | 0.002   | 0.000 | 0.002 | 0.007 | 0.013 | 0.019 | 0.020 | 0.007 |       |       |       |       |
| 84      | 0.002   | 0.001 | 0.003 | 0.010 | 0.016 | 0.029 | 0.028 | 0.013 | 0.002 |       |       |       |
| 96      | 0.002   | 0.001 | 0.003 | 0.011 | 0.018 | 0.029 | 0.027 | 0.015 | 0.003 | 0.000 |       |       |
| 108     | 0.002   | 0.001 | 0.004 | 0.011 | 0.019 | 0.031 | 0.027 | 0.014 | 0.003 | 0.001 | 0.000 |       |
| 120     | 0.003   | 0.001 | 0.004 | 0.012 | 0.021 | 0.033 | 0.030 | 0.014 | 0.005 | 0.003 | 0.002 | 0.000 |

To induce a risk averse representative investor to hold the market portfolio in Merton's (1973) ICAPM world, it must promise a higher excess return as the market's conditional variance increases. For an investor endowed with isoelastic preferences, the intensity of this risk-return relation is measured by the slope coefficient which corresponds to her relative risk aversion. Zero-intercept regressions are consistent with Merton's (1973) ICAPM expectations under the assumption that market excess return and past market variance are good proxies for conditional market risk premium and variance, respectively. Campbell and Thompson (2008) also advocate for this restriction in the ICAPM estimation. Our findings suggest that the conditional expected excess return on the stock market moves positively and proportionally with the market's conditional variance under the zero-intercept restriction. Hence, the risk-return term structure appears flat across the holding periods we consider in our empirical analysis. This points towards the constancy of the relative risk aversion of a representative investor with power utility looking at various investment horizons.

#### 4.4 Smoothed Estimator

The empirical risk-return trade-off analysis can be performed at different horizons. Yet, there is not much consensus on which data frequency to choose. This may explain some conflicting evidence found in the finance literature. Nonetheless, when the effect of extreme variance observations in the sample is dampened

through windsorization, we fail to reject the null hypothesis of a flat term structure of risk-return trade-offs for zero-intercept estimations. This motivates the idea of computing a single estimate for the risk-return trade-off intensity which will summarize the whole term structure. Interestingly, model averaging methods can combine our multi-horizon set of slopes into one cross-horizon estimate. One can view this single constrained estimator as a “backward-looking” proxy for the market price of risk which is robust to the horizon effect.

We now introduce a general framework for model averaging. Let  $D$  be the available dataset and  $\mathcal{M}_h$  a model specified at a given horizon  $h \in \mathcal{H} = \{1, \dots, 120\}$ . Actually,  $\mathcal{H}$  determines the number of all possible models in the analysis. Typically, this set can be very large. Given that the “true” specification is unknown, a single model selection is less appealing as it amounts to neglecting some information in the  $\text{card}(\mathcal{H}) - 1$  remaining models. A valuable alternative is to combine the parameter estimates from the different models, weighting them with their corresponding posterior probabilities  $\{p(\mathcal{M}_h | D)\}_{h \in \mathcal{H}}$ . Following formulas from Leamer (1978), the posterior mean and variance of the parameter in the composite model are

$$E(\beta | D) = \sum_{h \in \mathcal{H}} p(\mathcal{M}_h | D) \times E(\beta | \mathcal{M}_h, D), \quad (10)$$

and

$$\text{Var}(\beta | D) = \sum_{h \in \mathcal{H}} p(\mathcal{M}_h | D) \left[ \text{Var}(\beta | \mathcal{M}_h, D) + (E(\beta | \mathcal{M}_h, D) - E(\beta | D))^2 \right]. \quad (11)$$

The posterior variance formula illustrates both estimation and model risks. The first term in the posterior variance expression reflects the aggregation of model-specific estimation risks. The second term captures the variability across the models and therefore, it measures model uncertainty.

In the frequentist framework, the model averaging methodology proposes a smoothed estimator (see, e.g., Sala-i-Martin et al., 2004 and Hansen, 2007)

$$\hat{\beta}_{sm} = \sum_{h \in \mathcal{H}} \omega_h \times \hat{\beta}_h. \quad (12)$$

Note that for a given set of variables,  $E(\beta | \mathcal{M}_h, D)$  coincides with the least squares coefficient  $\hat{\beta}_h$  assum-

ing normality. Moreover, the weight  $\omega_h$  can be obtained from estimation criteria, including Akaike information criterion, Mallows criterion, Bayesian information criterion, etc. Based on the Akaike information criterion, the corresponding weight for Model  $\mathcal{M}_h$  is  $\omega_h = \exp(-\frac{1}{2}AIC_{\mathcal{M}_h}) / \sum_{l \in \mathcal{H}} \exp(-\frac{1}{2}AIC_{\mathcal{M}_l})$ . Similarly, the Bayesian information criterion weight of Model  $\mathcal{M}_h$  is  $\omega_h = \exp(-\frac{1}{2}BIC_{\mathcal{M}_h}) / \sum_{l \in \mathcal{H}} \exp(-\frac{1}{2}BIC_{\mathcal{M}_l})$ . The Mallows criterion-based weight vector  $W = (\omega_1, \dots, \omega_{card(\mathcal{H})})'$  minimizes a linear-quadratic objective function. Thus, the empirical Mallows weight vector  $\widehat{W} = \arg \min_W W' \bar{e}' \bar{e} W + 2\sigma^2 K' W$  is calculated numerically since no closed-form solution is available.

Accordingly, the variance of the smoothed estimator is computed as

$$Var(\widehat{\beta}_{sm}) = \sum_{h \in \mathcal{H}} \omega_h \left[ Var(\widehat{\beta}_h) + (\widehat{\beta}_h - \widehat{\beta}_{sm})^2 \right]. \quad (13)$$

Hansen (2007) conducts some simulation experiments to gauge the relative performance of least squares model average estimators. He finds that the smoothed AIC and Mallows estimators achieve lower risk than the smoothed BIC estimator in most simulation scenarios.

**Table 8**  
Smoothed cross-horizon risk-return trade-off estimates.

|                            | AIC   | BIC   | Mallows |
|----------------------------|-------|-------|---------|
| $\widehat{\beta}_{sm}$     | 2.370 | 2.368 | 2.024   |
| $SE(\widehat{\beta}_{sm})$ | 0.535 | 0.538 | 0.767   |

Table 8 shows three cross-horizon estimates of the risk-return trade-off based on different information criteria. Their values are between 2 and 3, consistently with previous empirical estimations of the relative risk aversion at the market index level. Recently in Bali and Engle (2010), the risk aversion coefficient was estimated to be between 2 and 4 across several assets in a panel regression under a Dynamic Conditional Correlations.

## 5. Conclusion

Our empirical evidence suggests that the term structure of risk-return trade-offs is nearly flat for low frequency data (monthly horizon and beyond). We argue that the weak evidence of risk-return trade-off at

the short horizon can be explained by sampling variation. For zero-intercept regressions, the linkages between market excess returns and past market variances appear both statistically and economically significant at all horizons. Based on the *rescaled* Wald statistics, we find that the differences between the estimated slope coefficients are statistically insignificant across horizons. It is worth noting that the zero-intercept restriction in the risk-return analysis is intended to yield more precise slope estimates as pointed out by Lanne and Saikkonen (2006), BP (2008), Campbell and Thompson (2008). A flat term structure of risk-return trade-offs, as suggested by our empirical findings, is broadly consistent with the implications of Merton-type ICAPM. Specifically, our empirical evidence does not indicate that different investment holding periods can generate notable changes in the risk-return trade-off intensity at the market level. Furthermore, model averaging techniques allow us to compute cross-horizon estimates of the risk-return trade-offs. The risk-return trade-off intensity is often considered as a proxy for the risk aversion of a representative investor. Our results show that the cross-horizon proxies for the relative risk aversion coefficient of an aggregate investor lie in a reasonable range, between 2 and 3.

We prove the econometric validity of our approach based on robust theoretical and empirical arguments. We rely on the asymptotic joint distribution of the set of trade-off intensities to provide some analytical inference results. We also implement some simulations to support the theoretical derivations. Our methodology rigorously incorporates the persistence in the realized variance using short- and long-memory dynamics.

An important research avenue is to study how risk-return models translate into optimal allocations from a long term investment perspective. Clearly, a strategic investor should care about both estimation and misspecification risks in choosing her optimal portfolio. We leave this topic for future research.

## Acknowledgements

The authors have benefited from stimulating comments of the editor (Erricos Kontoghiorghes), an associate editor, and two anonymous referees. The authors also thank Olivier Scaillet, Georges Dionne, Benoit Perron, Andrea Tamoni, Prosper Dovonon, Natalia Sizova, Frederico Bandi, Bryan Campbell and participants at the 2013 Conference on Computational and Financial Econometrics, 2013 CIREQ Econometrics Conference, the 2013 Mathematical Finance Days, the 2013 CIRPÉE Applied Financial Time Series Workshop and the 2011

Canadian Economic Association Meeting for helpful comments and discussion. The work of Cédric Okou is supported by the Institut de Finance Mathématique de Montréal, the Institut de la Finance Structurée et des Instruments Dérivés de Montréal and the UQAM research fund. Éric Jacquier acknowledges support from the HEC professorship in derivative securities and the MIT Sloan research fund.

## Appendix A. Proofs

A sketch for the proof of the propositions and corollaries proceeds in two main steps. First, the statistics of interest are properly scaled and re-expressed in terms of the predictor and predicted variables of the long-run risk-return regressions. Second, the asymptotic results are derived by invoking the Functional Central Limit Theorem (FCLT) and the Continuous Mapping Theorem (CMT).

**Proof of Proposition 1.** Assumptions (1-4) imply the following convergence result for a generic  $h_l$ -horizon regression slope estimate:

$$T\hat{\beta}_{h_l} = \frac{\frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{R_{t,t+h_l} - \bar{R}_{h_l}}{T^{1/2}} \right) \left( \frac{RV_{t-h_l,t} - \bar{RV}_{-h_l}}{T^{3/2}} \right)}{\frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l,t} - \bar{RV}_{-h_l}}{T^{3/2}} \right)^2} \xrightarrow{\mathcal{L}} \frac{\sigma_\varepsilon}{\sigma_u} \frac{\int_{\lambda_l}^{1-\lambda_l} \bar{W}(s, \lambda_l) \bar{J}_c(s, -\lambda_l) ds}{\int_{\lambda_l}^{1-\lambda_l} \bar{J}_c^2(s, -\lambda_l) ds}, \quad l=i, j.$$

Under the same set of Assumptions (1-4), a scaled covariance matrix of slope coefficients estimated at two arbitrary horizons  $h_i$  and  $h_j$  (w.l.o.g.) is defined as:

$$T^3 \text{Var} \begin{pmatrix} \hat{\beta}_{h_i} \\ \hat{\beta}_{h_j} \end{pmatrix} = T^3 \begin{pmatrix} \text{Var}(\hat{\beta}_{h_i}) & \text{Cov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) \\ \text{Cov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) & \text{Var}(\hat{\beta}_{h_j}) \end{pmatrix}.$$

By rewriting the elements of the covariance matrix in terms of the regressor  $(RV_{t-h_l,t})$  and regressand  $(R_{t,t+h_l})$ , we obtain:

$$T^3 \text{Var}(\hat{\beta}_{h_l}) = \frac{S_l^2}{T} \left( \frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l,t} - \bar{RV}_{-h_l}}{T^{3/2}} \right)^2 \right)^{-1},$$

$$\text{where } S_l^2 = \frac{1}{T-2h_l+1} \left[ \sum_{t=1}^{T-2h_l+1} (R_{t,t+h_l} - \bar{R}_{h_l})^2 - \hat{\beta}_{h_l}^2 \sum_{t=1}^{T-2h_l+1} (RV_{t-h_l,t} - \bar{RV}_{-h_l})^2 \right], \quad l=i, j.$$

We also have:



$$T^3 Cov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) = \frac{S_{ij}}{T} \left[ \frac{\left( \frac{1}{T} \sum_{t=1}^{T-2h_{(i \vee j)}+1} \left( \frac{RV_{t-h_i, t} - \bar{RV}_{-h_i}}{T^{3/2}} \right) \left( \frac{RV_{t-h_j, t} - \bar{RV}_{-h_j}}{T^{3/2}} \right) \right)}{\left( \frac{1}{T} \sum_{t=1}^{T-2h_i+1} \left( \frac{RV_{t-h_i, t} - \bar{RV}_{-h_i}}{T^{3/2}} \right)^2 \right) \left( \frac{1}{T} \sum_{t=1}^{T-2h_j+1} \left( \frac{RV_{t-h_j, t} - \bar{RV}_{-h_j}}{T^{3/2}} \right)^2 \right)} \right],$$

where  $S_{ij} = \frac{1}{T-2h_{(i \vee j)}+1} \left[ \begin{array}{cc} \sum_{t=1}^{T-2h_{(i \vee j)}+1} (R_{t, t+h_i} - \bar{R}_{h_i}) (R_{t, t+h_j} - \bar{R}_{h_j}) & \\ -\hat{\beta}_{h_i} \hat{\beta}_{h_j} & \sum_{t=1}^{T-2h_{(i \vee j)}+1} (RV_{t-h_i, t} - \bar{RV}_{-h_i}) (RV_{t-h_j, t} - \bar{RV}_{-h_j}) \end{array} \right]$  is the covariance between the error terms of the two regressions, and  $h_{(i \vee j)} \equiv \max(h_i, h_j)$ .

Note that  $\bar{R}_{h_l} = \frac{1}{T-2h_l+1} \sum_{t=1}^{T-2h_l+1} R_{t, t+h_l}$  and  $\bar{RV}_{-h_l} = \frac{1}{T-2h_l+1} \sum_{t=1}^{T-2h_l+1} RV_{t-h_l, t}$ ,  $l = i, j$ , are sample averages of the regressand and regressor for a generic  $h_l$ -horizon predictive risk-return regression.

Now, by invoking the FCLT and the CMT it follows that:

$$\frac{S_l^2}{T} \left( \frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l, t} - \bar{RV}_{-h_l}}{T^{3/2}} \right)^2 \right)^{-1} = \frac{T}{T-2h_l+1} \frac{\left[ \frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{R_{t, t+h_l} - \bar{R}_{h_l}}{T^{1/2}} \right)^2 - \left( T \hat{\beta}_{h_l} \right)^2 \frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l, t} - \bar{RV}_{-h_l}}{T^{3/2}} \right)^2 \right]}{\frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l, t} - \bar{RV}_{-h_l}}{T^{3/2}} \right)^2}$$

$$\xrightarrow{\mathcal{L}} \frac{\sigma_u^2}{\sigma_u^2} \frac{\left[ \int_{\lambda_l}^{1-\lambda_l} \bar{W}^2(s, \lambda_l) ds \int_{\lambda_l}^{1-\lambda_l} \bar{J}_c^2(s, -\lambda_l) ds - \left( \int_{\lambda_l}^{1-\lambda_l} \bar{W}(s, \lambda_l) \bar{J}_c(s, -\lambda_l) ds \right)^2 \right]}{(1-2\lambda_l) \left( \int_{\lambda_l}^{1-\lambda_l} \bar{J}_c^2(s, -\lambda_l) ds \right)^2}, l = i, j.$$

It also holds that:

$$\frac{S_{ij}}{T} \left[ \frac{\left( \frac{1}{T} \sum_{t=1}^{T-2h_{(i \vee j)}+1} \left( \frac{RV_{t-h_i, t} - \bar{RV}_{-h_i}}{T^{3/2}} \right) \left( \frac{RV_{t-h_j, t} - \bar{RV}_{-h_j}}{T^{3/2}} \right) \right)}{\left( \frac{1}{T} \sum_{t=1}^{T-2h_i+1} \left( \frac{RV_{t-h_i, t} - \bar{RV}_{-h_i}}{T^{3/2}} \right)^2 \right) \left( \frac{1}{T} \sum_{t=1}^{T-2h_j+1} \left( \frac{RV_{t-h_j, t} - \bar{RV}_{-h_j}}{T^{3/2}} \right)^2 \right)} \right]$$

$$\xrightarrow{\mathcal{L}} \frac{\sigma_u^2}{\sigma_u^2} \frac{1}{1-2\lambda_{(i \vee j)}} \left[ \frac{\int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \bar{W}(s, \lambda_i) \bar{W}(s, \lambda_j) ds \int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \bar{J}_c(s, -\lambda_i) \bar{J}_c(s, -\lambda_j) ds}{\int_{\lambda_i}^{1-\lambda_i} \bar{J}_c^2(s, -\lambda_i) ds \int_{\lambda_j}^{1-\lambda_j} \bar{J}_c^2(s, -\lambda_j) ds} - \frac{\left( \int_{\lambda_i}^{1-\lambda_i} \bar{W}(s, \lambda_i) \bar{J}_c(s, -\lambda_i) ds \int_{\lambda_j}^{1-\lambda_j} \bar{W}(s, \lambda_j) \bar{J}_c(s, -\lambda_j) ds \right)^2}{\left( \int_{\lambda_i}^{1-\lambda_i} \bar{J}_c^2(s, -\lambda_i) ds \right)^2 \left( \int_{\lambda_j}^{1-\lambda_j} \bar{J}_c^2(s, -\lambda_j) ds \right)^2} \right].$$

$$\text{Hence, } T^3 Var(\hat{\beta}_{h_l}) \xrightarrow{\mathcal{L}} \overline{Avar}(\hat{\beta}_{h_l}) = \frac{\sigma_u^2}{\sigma_u^2} \frac{\left[ \int_{\lambda_l}^{1-\lambda_l} \bar{W}^2(s, \lambda_l) ds \int_{\lambda_l}^{1-\lambda_l} \bar{J}_c^2(s, -\lambda_l) ds - \left( \int_{\lambda_l}^{1-\lambda_l} \bar{W}(s, \lambda_l) \bar{J}_c(s, -\lambda_l) ds \right)^2 \right]}{(1-2\lambda_l) \left( \int_{\lambda_l}^{1-\lambda_l} \bar{J}_c^2(s, -\lambda_l) ds \right)^2}, l=i, j.$$

Finally,  $T^3 Cov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) \xrightarrow{\mathcal{L}} \overline{Acov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j})$ , where

$$\overline{Acov} \left( \hat{\beta}_{h_i}, \hat{\beta}_{h_j} \right) = \frac{\sigma_\varepsilon^2}{\sigma_u^2 (1 - 2\lambda_{(i \vee j)})} \left[ \begin{array}{c} \frac{\int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \overline{W}(s, \lambda_i) \overline{W}(s, \lambda_j) ds \int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \overline{J}_c(s, -\lambda_i) \overline{J}_c(s, -\lambda_j) ds}{\int_{\lambda_i}^{1-\lambda_i} \overline{J}_c^2(s, -\lambda_i) ds \int_{\lambda_j}^{1-\lambda_j} \overline{J}_c^2(s, -\lambda_j) ds} \\ - \frac{\int_{\lambda_i}^{1-\lambda_i} \overline{W}(s, \lambda_i) \overline{J}_c(s, -\lambda_i) ds \int_{\lambda_j}^{1-\lambda_j} \overline{W}(s, \lambda_j) \overline{J}_c(s, -\lambda_j) ds \left( \int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \overline{J}_c(s, -\lambda_i) \overline{J}_c(s, -\lambda_j) ds \right)^2}{\left( \int_{\lambda_i}^{1-\lambda_i} \overline{J}_c^2(s, -\lambda_i) ds \right)^2 \left( \int_{\lambda_j}^{1-\lambda_j} \overline{J}_c^2(s, -\lambda_j) ds \right)^2} \end{array} \right],$$

and  $\lambda_{(i \vee j)} \equiv \max(\lambda_i, \lambda_j)$ .

The proof of Proposition 2 is similar except that for the restricted regressions,  $\overline{W}(s, \lambda)$  and  $\overline{J}_c(s, -\lambda)$  are replaced with their uncentered versions  $W(s, \lambda)$  and  $J_c(s, -\lambda)$ .

**Proof of Proposition 3.** Under Assumptions (1-2, 5-6), the following convergence result for a generic  $h_l$ -horizon regression slope estimate holds:

$$T^d \hat{\beta}_{h_l} = \frac{\frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{R_{t,t+h_l} - \overline{R}_{h_l}}{T^{1/2}} \right) \left( \frac{RV_{t-h_l,t} - \overline{RV}_{-h_l}}{T^{1/2+d}} \right)}{\frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l,t} - \overline{RV}_{-h_l}}{T^{1/2+d}} \right)^2} \xrightarrow{\mathcal{L}} \frac{\sigma_0}{\sigma_d} \frac{\int_{\lambda_l}^{1-\lambda_l} \overline{W}(s, \lambda_l) \overline{B}_d(s, -\lambda_l) ds}{\int_{\lambda_l}^{1-\lambda_l} \overline{B}_d^2(s, -\lambda_l) ds}, \quad l=i,j.$$

Moreover, if Assumptions (1-2, 5-6) are met, we can scale a covariance matrix of the slope coefficients estimated at two arbitrary horizons  $h_i$  and  $h_j$  (w.l.o.g.) as follows:

$$T^{1+2d} Var \begin{pmatrix} \hat{\beta}_{h_i} \\ \hat{\beta}_{h_j} \end{pmatrix} = T^{1+2d} \begin{pmatrix} Var(\hat{\beta}_{h_i}) & Cov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) \\ Cov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) & Var(\hat{\beta}_{h_j}) \end{pmatrix}.$$

Thus, the diagonal elements of the covariance matrix may be re-expressed in terms of the regressor  $(RV_{t-h_l,t})$  and regressand  $(R_{t,t+h_l})$  as:

$$T^{1+2d} Var(\hat{\beta}_{h_l}) = \frac{S_l^2}{T} \left( \frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l,t} - \overline{RV}_{-h_l}}{T^{1/2+d}} \right)^2 \right)^{-1},$$

$$\text{where } S_l^2 = \frac{1}{T-2h_l+1} \left[ \sum_{t=1}^{T-2h_l+1} (R_{t,t+h_l} - \overline{R}_{h_l})^2 - \hat{\beta}_{h_l}^2 \sum_{t=1}^{T-2h_l+1} (RV_{t-h_l,t} - \overline{RV}_{-h_l})^2 \right], \quad l=i,j.$$

Similarly, we can rewrite the off-diagonal elements as:

$$T^{1+2d} Cov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) = \frac{S_{ij}}{T} \left[ \frac{\left( \frac{1}{T} \sum_{t=1}^{T-2h_{(i \vee j)}+1} \left( \frac{RV_{t-h_i,t} - \overline{RV}_{-h_i}}{T^{1/2+d}} \right) \left( \frac{RV_{t-h_j,t} - \overline{RV}_{-h_j}}{T^{1/2+d}} \right) \right)}{\left( \frac{1}{T} \sum_{t=1}^{T-2h_i+1} \left( \frac{RV_{t-h_i,t} - \overline{RV}_{-h_i}}{T^{1/2+d}} \right)^2 \right) \left( \frac{1}{T} \sum_{t=1}^{T-2h_j+1} \left( \frac{RV_{t-h_j,t} - \overline{RV}_{-h_j}}{T^{1/2+d}} \right)^2 \right)} \right],$$

where  $S_{ij} = \frac{1}{T-2h_{(i \vee j)}+1} \begin{bmatrix} \sum_{t=1}^{T-2h_{(i \vee j)}+1} (R_{t,t+h_i} - \bar{R}_{h_i}) (R_{t,t+h_j} - \bar{R}_{h_j}) \\ -\hat{\beta}_{h_i} \hat{\beta}_{h_j} \sum_{t=1}^{T-2h_{(i \vee j)}+1} (RV_{t-h_i,t} - \bar{RV}_{-h_i}) (RV_{t-h_j,t} - \bar{RV}_{-h_j}) \end{bmatrix}$  is the covariance between the error terms of the two regressions and  $h_{(i \vee j)} \equiv \max(h_i, h_j)$ .

The FCLT and the CMT allow us to compute the following convergence result:

$$\begin{aligned} \frac{S_l^2}{T} \left( \frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l,t} - \bar{RV}_{-h_l}}{T^{1/2+d}} \right)^2 \right)^{-1} &= \frac{1}{T-2h_l+1} \frac{\left[ \sum_{t=1}^{T-2h_l+1} \left( \frac{R_{t,t+h_l} - \bar{R}_{h_l}}{T^{1/2}} \right)^2 - (T^d \hat{\beta}_{h_l})^2 \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l,t} - \bar{RV}_{-h_l}}{T^{1/2+d}} \right)^2 \right]}{\frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l,t} - \bar{RV}_{-h_l}}{T^{1/2+d}} \right)^2} \\ &= \frac{T}{T-2h_l+1} \frac{\left[ \frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{R_{t,t+h_l} - \bar{R}_{h_l}}{T^{1/2}} \right)^2 - (T^d \hat{\beta}_{h_l})^2 \frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l,t} - \bar{RV}_{-h_l}}{T^{1/2+d}} \right)^2 \right]}{\frac{1}{T} \sum_{t=1}^{T-2h_l+1} \left( \frac{RV_{t-h_l,t} - \bar{RV}_{-h_l}}{T^{1/2+d}} \right)^2} \\ &\xrightarrow{\mathcal{L}} \frac{\sigma_{\vartheta}^2}{\sigma_d^2} \frac{\left[ \int_{\lambda_l}^{1-\lambda_l} \bar{W}^2(s, \lambda_l) ds \int_{\lambda_l}^{1-\lambda_l} \bar{B}_d^2(s, -\lambda_l) ds - \left( \int_{\lambda_l}^{1-\lambda_l} \bar{W}(s, \lambda_l) \bar{B}_d(s, -\lambda_l) ds \right)^2 \right]}{(1-2\lambda_l) \left( \int_{\lambda_l}^{1-\lambda_l} \bar{B}_d^2(s, -\lambda_l) ds \right)^2}, l = i, j. \end{aligned}$$

It also holds that:

$$\begin{aligned} \frac{S_{ij}}{T} \left[ \frac{\left( \frac{1}{T} \sum_{t=1}^{T-2h_{(i \vee j)}+1} \left( \frac{RV_{t-h_i,t} - \bar{RV}_{-h_i}}{T^{1/2+d}} \right) \left( \frac{RV_{t-h_j,t} - \bar{RV}_{-h_j}}{T^{1/2+d}} \right) \right)}{\left( \frac{1}{T} \sum_{t=1}^{T-2h_i+1} \left( \frac{RV_{t-h_i,t} - \bar{RV}_{-h_i}}{T^{1/2+d}} \right)^2 \right) \left( \frac{1}{T} \sum_{t=1}^{T-2h_j+1} \left( \frac{RV_{t-h_j,t} - \bar{RV}_{-h_j}}{T^{1/2+d}} \right)^2 \right)} \right] \\ \xrightarrow{\mathcal{L}} \frac{\sigma_{\vartheta}^2}{\sigma_d^2 (1-2\lambda_{(i \vee j)})} \left[ \frac{\frac{\int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \bar{W}(s, \lambda_i) \bar{W}(s, \lambda_j) ds \int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \bar{B}_d(s, -\lambda_i) \bar{B}_d(s, -\lambda_j) ds}{\int_{\lambda_i}^{1-\lambda_i} \bar{B}_d^2(s, -\lambda_i) ds \int_{\lambda_j}^{1-\lambda_j} \bar{B}_d^2(s, -\lambda_j) ds} \right. \\ \left. - \frac{\int_{\lambda_i}^{1-\lambda_i} \bar{W}(s, \lambda_i) \bar{B}_d(s, -\lambda_i) ds \int_{\lambda_j}^{1-\lambda_j} \bar{W}(s, \lambda_j) \bar{B}_d(s, -\lambda_j) ds \left( \int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \bar{B}_d(s, -\lambda_i) \bar{B}_d(s, -\lambda_j) ds \right)^2}{\left( \int_{\lambda_i}^{1-\lambda_i} \bar{B}_d^2(s, -\lambda_i) ds \right)^2 \left( \int_{\lambda_j}^{1-\lambda_j} \bar{B}_d^2(s, -\lambda_j) ds \right)^2} \right]. \end{aligned}$$

Hence,  $T^{1+2d} Var(\hat{\beta}_{h_l}) \xrightarrow{\mathcal{L}} \overline{Avar}(\hat{\beta}_{h_l}) = \frac{\sigma_{\vartheta}^2}{\sigma_d^2} \frac{\left[ \int_{\lambda_l}^{1-\lambda_l} \bar{W}^2(s, \lambda_l) ds \int_{\lambda_l}^{1-\lambda_l} \bar{B}_d^2(s, -\lambda_l) ds - \left( \int_{\lambda_l}^{1-\lambda_l} \bar{W}(s, \lambda_l) \bar{B}_d(s, -\lambda_l) ds \right)^2 \right]}{(1-2\lambda_l) \left( \int_{\lambda_l}^{1-\lambda_l} \bar{B}_d^2(s, -\lambda_l) ds \right)^2}, l = i, j.$

Finally,  $T^{1+2d} Cov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) \xrightarrow{\mathcal{L}} \overline{Acov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j})$ , where

$$\overline{Acov}(\hat{\beta}_{h_i}, \hat{\beta}_{h_j}) = \frac{\sigma_{\vartheta}^2}{\sigma_d^2 (1-2\lambda_{(i \vee j)})} \left[ \frac{\frac{\int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \bar{W}(s, \lambda_i) \bar{W}(s, \lambda_j) ds \int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \bar{B}_d(s, -\lambda_i) \bar{B}_d(s, -\lambda_j) ds}{\int_{\lambda_i}^{1-\lambda_i} \bar{B}_d^2(s, -\lambda_i) ds \int_{\lambda_j}^{1-\lambda_j} \bar{B}_d^2(s, -\lambda_j) ds} \right. \\ \left. - \frac{\int_{\lambda_i}^{1-\lambda_i} \bar{W}(s, \lambda_i) \bar{B}_d(s, -\lambda_i) ds \int_{\lambda_j}^{1-\lambda_j} \bar{W}(s, \lambda_j) \bar{B}_d(s, -\lambda_j) ds \left( \int_{\lambda_{(i \vee j)}}^{1-\lambda_{(i \vee j)}} \bar{B}_d(s, -\lambda_i) \bar{B}_d(s, -\lambda_j) ds \right)^2}{\left( \int_{\lambda_i}^{1-\lambda_i} \bar{B}_d^2(s, -\lambda_i) ds \right)^2 \left( \int_{\lambda_j}^{1-\lambda_j} \bar{B}_d^2(s, -\lambda_j) ds \right)^2} \right],$$

and  $\lambda_{(i \vee j)} \equiv \max(\lambda_i, \lambda_j)$ .

The proof of Proposition 4 follows the same steps except that for the restricted regressions,  $\overline{W}(s, \lambda)$  and  $\overline{B}_d(s, -\lambda)$  are replaced with their uncentered counterparts  $W(s, \lambda)$  and  $B_d(s, -\lambda)$ .

**Proof of Corollary 1.** Let  $\lambda_{(a \vee b)}$  denote the maximum between two distinct fractions of overlap  $\lambda_a$  and  $\lambda_b$ . To derive the asymptotic distribution of a *rescaled*-Wald statistic, we consider the following two four-argument functionals of stochastic processes that are defined and almost surely continuous over  $[0, 1]$ :

$$\mathcal{N}(X_s(\lambda_a), X_s(\lambda_b), Y_s(\lambda_a), Y_s(\lambda_b)) \equiv \{F_1(X_s(\lambda_b), Y_s(\lambda_b)) - F_1(X_s(\lambda_a), Y_s(\lambda_a))\}^2,$$

and

$$\begin{aligned} \mathcal{D}(X_s(\lambda_a), X_s(\lambda_b), Y_s(\lambda_a), Y_s(\lambda_b)) &\equiv \left\{ \frac{F_2(X_s(\lambda_a), Y_s(\lambda_a))}{1 - 2\lambda_a} \right. \\ &\quad + \frac{F_2(X_s(\lambda_b), Y_s(\lambda_b))}{1 - 2\lambda_b} \\ &\quad \left. - \frac{2 \times F_3(X_s(\lambda_a), X_s(\lambda_b), Y_s(\lambda_a), Y_s(\lambda_b))}{1 - 2\lambda_{(a \vee b)}} \right\}, \end{aligned}$$

where  $F_1(\bullet)$ ,  $F_2(\bullet)$ , and  $F_3(\bullet)$  are given in the main text.

To investigate the horizon effect in the risk-return relationship estimated at two distinct horizons  $h_i$  and  $h_j$  (w.l.o.g.), we use a *rescaled* Wald statistic and test the following restriction  $\mathcal{G} \begin{bmatrix} \hat{\beta}_{h_i} \\ \hat{\beta}_{h_j} \end{bmatrix} = 0$  where  $\mathcal{G} = \begin{bmatrix} -1 & 1 \end{bmatrix}$ .

$$\text{Recall that } T^{-1} \chi^2_{\left( \hat{\beta}_{h_j} = \lfloor \lambda_j T \rfloor - \hat{\beta}_{h_i} = \lfloor \lambda_i T \rfloor \right)} = \frac{[T(\hat{\beta}_{h_j} - \hat{\beta}_{h_i} - 0)]^2}{T^3 [Var(\hat{\beta}_{h_i}) + Var(\hat{\beta}_{h_j}) - 2Cov(\hat{\beta}_{h_i}, \hat{\beta}_{h_j})]}.$$

Hence, combining the FCLT, the CMT and the results of Propositions (1) and (2) yields:

for the unrestricted regressions,

$$T^{-1} \chi^2_{\left( \hat{\beta}_{h_j} = \lfloor \lambda_j T \rfloor - \hat{\beta}_{h_i} = \lfloor \lambda_i T \rfloor \right)} \xrightarrow{\mathcal{L}} \frac{\mathcal{N}(\overline{W}(s, \lambda_i), \overline{W}(s, \lambda_j), \overline{J}_c(s, -\lambda_i), \overline{J}_c(s, -\lambda_j))}{\mathcal{D}(\overline{W}(s, \lambda_i), \overline{W}(s, \lambda_j), \overline{J}_c(s, -\lambda_i), \overline{J}_c(s, -\lambda_j))}.$$

and for the restricted regressions,

$$T^{-1}\chi^2_{\left(\hat{\beta}_{h_j=\lfloor \lambda_j T \rfloor} - \hat{\beta}_{h_i=\lfloor \lambda_i T \rfloor}\right)} \xrightarrow{\mathcal{L}} \frac{\mathcal{N}(W(s, \lambda_i), W(s, \lambda_j), J_c(s, -\lambda_i), J_c(s, -\lambda_j))}{\mathcal{D}(W(s, \lambda_i), W(s, \lambda_j), J_c(s, -\lambda_i), J_c(s, -\lambda_j))}.$$

This completes the proof of Corollary 1.

The proof of Corollary 2 proceeds similarly, by invoking the FCLT and the CMT and using the convergence results of Propositions (3) and (4).

## References

- Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. *Journal of Finance* 47, 1047-1092.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., 2010. Parametric and Nonparametric Volatility Measurement. In: Y. Ait-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, North-Holland, 67-138.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71, 579-625.
- Ang, A., Bekaert, G., 2007. Stock Return Predictability: Is it There? *Review of Financial Studies* 20, 651-707.
- Bali, T.G., Engle, R.T., 2010. The Intertemporal Capital Asset Pricing Model with Dynamic Conditional Correlations. *Journal of Monetary Economics* 57, 377-390.
- Baillie, R.T., 1996. Long Memory Processes and Fractional Integration in Econometrics. *Journal of Econometrics* 73, 5-59.
- Bandi, F.M., Perron, B., 2006. Long Memory and the Relation Between Realized and Implied Volatility. *Journal of Financial Econometrics* 4, 636-670.
- Bandi, F.M., Perron, B., 2008. Long-Run Risk>Returns Trade-Offs. *Journal of Econometrics* 143, 349-374.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481-1536.
- Barndorff-Nielsen, O.E., Shephard, N., 2003. Realised Power Variation and Stochastic Volatility. *Bernoulli* 9, 243-265.
- Beckers, S., 1983. Variances of security price returns based on high, low, and closing prices. *Journal of Business* 56, 97-112.
- Beran, J., 1994. *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

- Bollerslev, T., Tauchen, G., Zhou, H. 2009. Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies* 22, 4463–4492.
- Bollerslev, T., Sizova, N., Tauchen, G., 2012. Volatility in Equilibrium: Asymmetries and Dynamic Dependencies. *Review of Finance* 16, 31–81.
- Boudoukh, J., Richardson, M., 1993. Stock Returns and Inflation: A Long-Horizon Perspective. *American Economic Review* 83, 1346–1355.
- Boudoukh, J., Richardson, M., Whitelaw, R.F., 2008. The Myth of Long-Horizon Predictability. *Review of Financial Studies* 21, 1577–1605.
- Breidt, F.J., Crato, N., de Lima, P., 1994. The Detection and Estimation of Long Memory in Stochastic Volatility. *Journal of Econometrics* 83, 325–348.
- Campbell, J.Y., 1987. Stock Returns and the Term Structure. *Journal of Financial Economics* 18, 373–399.
- Campbell, J.Y., 2001. Why Long Horizons? A Study of Power Against Persistent Alternatives. *Journal of Empirical Finance* 8, 459–491.
- Campbell, J.Y., Thompson, S.B., 2008. Predicting the Equity Premium out of Sample: Can Anything Beat the Historical Average. *Review of Financial Studies* 21, 1509–1531.
- Campbell, J.Y., Shiller, R., 1988. The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies* 1, 195–227.
- Campbell, J.Y., and Viceira, L., 2005. The Term Structure of the Risk-Return Tradeoff. NBER Working Papers 11119, National Bureau of Economic Research.
- Campbell, J.Y., and Yogo, M. 2006. Efficient Tests of Stock Return Predictability. *Journal of Financial Economics* 81, 27–60.
- Camponovo, L., Scaillet, O., Trojani, F., 2013. Predictability Hidden by Anomalous Observations. Working Paper HEC Genève.
- Comte, F., Renault, E., 1998. Long Memory in Continuous-Time Stochastic Volatility Models. *Mathematical Finance* 8, 291–323.
- Corsi, F., 2009. A Simple Approximate Long-Memory model of Realized Volatility. *Journal of Financial Econometrics* 7, 174–196.
- Christensen, K., Podolskij, M., 2007. Realized range-based estimation of integrated variance. *Journal of Econometrics* 141, 323–349.
- Davidson, J., and Hashimzade, N., 2009. Type I and Type II Fractional Brownian Motions: a Reconsideration. *Computational Statistics and Data Analysis* 53, 2089–2106.
- DeGennaro, R.P., Zhao, Y.L., 1998. Stock Returns and Volatility: Another Look. *Journal of Economics and Finance* 22, 5–18.
- Diebold, F.X., Inoue, A., 2001. Long memory and regime switching. *Journal of Econometrics* 105, 131–159.

- Dijk, D.J.C.V., Martens, M., 2007. Measuring volatility with the realized range. *Journal of Econometrics* 138, 181–207.
- Fama, E.F., French, K.R., 1988. Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 22, 3–25.
- French, K.R., Schwert, W., Stambaugh, R.F., 1987. Expected Stock Returns and Volatility. *Journal of Financial Economics* 19, 3–29.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There Is a Risk-Return Trade-Off After All. *Journal of Financial Economics* 76, 509–548.
- Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance* 48, 1779–1801.
- Goyal, A., Welch, I., 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Predictions. *Review of Financial Studies* 21, 1455–1508.
- Guo, H., Whitelaw, R., 2006. Uncovering the Risk-Return Relation in the Stock Market. *Journal of Finance* 61, 1433–1464.
- Hansen, B.E., 2007. Least Squares Model Averaging. *Econometrica* 75, 1175–1189.
- Hodrick, R., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. *Review of Financial Studies* 5, 357–386.
- Jacquier, E., Okou, C., 2013. Disentangling Continuous Volatility from Jumps in Long-Run Risk-Return Relationships. *Journal of Financial Econometrics*, Forthcoming.
- Jacquier, E., Polson, N.G., Rossi, P.E., 2004. Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors. *Journal of Econometrics* 122, 185–212.
- Lanne, M., Saikkonen, P., 2006. Why Is It so Difficult to Uncover the Risk-Return Trade-Off in Stock Returns? *Economics Letters* 1, 118–125.
- Leamer, E.E., 1978. *Specification Searches*. New York: John Wiley & Sons.
- Lettau, M., Ludvigson, S.C., 2001. Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance* 56, 815–849.
- Lettau, M., Ludvigson, S.C., 2010. Measuring and Modeling Variation in the Risk-Return Trade-Off. In: Y. Ait-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, North-Holland, 617–690.
- Maheu, J., McCurdy, T., 2007. Components of Market Risk and Return. *Journal of Financial Econometrics* 5, 560–590.
- Mandelbrot, B.B., Van Ness, J.W., 1968. Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review* 10, 422–437.
- Marinucci, D., Robinson, P.M., 2000. Weak Convergence of Multivariate Fractional Processes. *Stochastic Processes and their Applications* 86, 103–120.

- Merton, R.C., 1973. An Intertemporal Capital Asset Pricing Model. *Econometrica* 41, 867–887.
- Merton, R.C., 1980. On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics* 8, 323–361.
- Nelson, C., Kim, M., 1993. Predictable stock returns: the role of small sample bias. *Journal of Finance* 48, 641–661.
- Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. *Journal of Business* 53, 61–65.
- Sala-i-Martin, X., Doppelhofer, G., Miller, R., 2004. Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review* 94, 813–835.
- Sizova, N., 2013. Long-Horizon Return Regressions with Historical Volatility and Other Long-Memory Variables. *Journal of Business and Economic Statistics* 31, 546–559.
- Stambaugh, R.F., 1999. Predictive Regressions. *Journal of Financial Economics* 54 , 375–421.
- Taqqu, M.S., 1975. Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 31, 287–302.
- Tsay, W.J., Chung, C.F., 2000. The Spurious Regression of Fractionally Integrated Processes. *Journal of Econometrics* 96, 155–182.
- Valkanov, R., 2003. Long-horizon Regressions: Theoretical Results and Applications. *Journal of Financial Economics* 68, 201–232.