THE SMALL SAMPLE PROPERTIES OF TESTS OF THE EXPECTATIONS HYPOTHESIS: A MONTE CARLO INVESTIGATION

E. GARGANAS, S.G.HALL

Tanaka Business School Discussion Papers: TBS/DP04/6
London: Tanaka Business School, 2004
The small sample properties of tests of the expectations hypothesis:
A Monte Carlo investigation

By
E. Garganas (National Bank of Greece), S.G. Hall (Imperial College)

ABSTRACT
In this paper, we extend results from the finance literature that explores small-sample bias, due to persistent variables, in tests of present value asset pricing models. Using Monte Carlo experiments we analyze the finite sample behaviour of a variety of tests of the expectations hypothesis of the term structure, when interest rates are increasingly persistent. Results overwhelmingly indicate that asymptotic inference can be misleading. In the Campbell and Shiller (1991) vector autoregression (VAR) framework, the Wald test overrejects the correct null and the slope of the corresponding ordinary least squares (OLS) test is strongly biased. The distortions can be extreme for recursive estimators, such as, the Kalman filter time varying parameter model.

October 2004

**JEL classification:** C15; E43; G12;
**Keywords:** Interest rates; Expectations hypothesis; Small sample bias; Kalman filter
1. Introduction

The inconclusive empirical evidence against the expectations hypothesis of the term structure of interest rates has often led researchers to suggest that the theory remains popular primarily on *a priori* grounds, as it still seems to provide the most intuitive explanation for the term structure. In its ‘pure form’ the expectations hypothesis (EH) states that the long rate is a forecast of expected future short rates, assuming zero or a time-invariant risk premium and rational expectations. At the same time, these empirical puzzles have brought about the development of a small variety of interesting explanations for the theory’s rejection. A recurring issue in the recent empirical literature, has been the adoption of a formal methodology for excluding the unfortunate case in which the EH is merely rejected as a result of statistical or econometric distortions in small samples (and correcting for them). The particular case of strong small sample bias due to persistent interest rates is a typical problem that applies to many markets, including that of the U.S.

Generally, there are three broad explanations for empirical failures of the EH. First, the theory simply isn’t true. The key behavioural assumptions do not hold e.g. agents are irrational or risk averse. This has turned a relatively small part of research away from US interest rates for which the EH is usually rejected (e.g. Campbell and Shiller, 1991; Bekaert et al., 1997) towards other, mainly European, market rates for several of which the EH is not rejected (e.g., Hardouvelis, 1994; Gerlach and Smets, 1997; Bekaert and Hodrick, 2000). Second, the EH can be ‘saved’ by incorporating time-varying risk premia to the model specification (GARCH models) (Engle, et al., 1987). Third, the statistical tests themselves lead to false rejections due to their poor small sample properties. This leads to a family of explanations including peso problems with regime shifts (e.g. Bekaert et al., 2001), learning behaviour and, finally, persistent variables (Mankiw and Shapiro, 1986; Bekaert et al., 1997). The latter is the explanation that we set out to explore in this paper, using artificial rather than real world interest rates, in a Monte Carlo simulation environment.
In an influential paper, Bekaert et al. (1997) show that the small sample properties of standard tests of the EH can be surprisingly ‘twisted’ when interest rates are highly persistent. In their results, the two well known Campbell and Shiller (1991) ‘spread’ regression specifications and the two VAR methodology regressions, all provide strongly biased coefficients that depend on the persistence parameter of an AR(1) one-period rate, even for samples of 2000 observations. The high dispersion of the small sample distribution of the slope coefficient, makes it more ‘difficult to reject’ the EH in small samples. They conclude that small sample inference using carefully designed Monte Carlo simulation experiments has essentially become a ‘sine qua non’ for testing the EH in modern finance research (e.g. Longstaff, 2000).

In this paper, we conduct a series of Monte Carlo experiments to examine the small sample properties of the ‘perfect foresight spread’ regression test, the variance bounds inequality and four ‘VAR methodology’ statistics (Campbell and Shiller, 1991), all in response to an increasing persistence parameter. Given that the literature is not always conclusive on the properties of these small sample distributions, depending on the data generating assumptions (or the theoretical approximations in closed form estimates), we confirm that for most of these tests high persistence leads to a bias towards overrejection of the EH when asymptotic inference is used. The surprising exception is the variance bounds ratio for which the EH is underrejected. We also find a particularly strong bias in the case of the Wald test of the VAR coefficient restrictions and in the slope coefficient of the regression of the (VAR-constructed) ‘theoretical’ against the actual spread. Considering the further issue of interest rate maturity horizon we find that it matters significantly, for all tests. Overall, our Monte Carlo results indicate that the statistical size distortions are larger for some tests than for others. This would help explain why it is often the case that in actual markets different tests may be more favourable or less favourable to the EH.

The paper is organised as follows. Section 2 briefly outlines tests of the EH model. Section 3 explains why we may expect small sample bias to be particularly strong for most of these tests when interest rates are highly persistent. We discuss the literature findings and focus on the properties of the slope coefficient in the Campbell and Shiller ‘perfect foresight spread’ regression. Section 4 explores Monte Carlo evidence regarding the small sample behaviour of 6 standard EH tests and a Kalman filter model. A sensitivity analysis is performed with respect to a changing persistence coefficient in an AR(1) data generating process for the short rate. The last section then summarises the key findings with some closing remarks.
2. Empirical tests of the EH

2.1 The ‘perfect foresight spread’ regression and volatility tests

The main term structure relationship implied by the EH using continuously compounded interest rates and a logarithmic approximation for pure discount bonds is:

\[
R^n_t = \left( \frac{1}{n} \right) \sum_{i=0}^{n-1} E_t r_{i+1} + c
\]  

(1)

where the short rate, with one period to maturity, is denoted by \( r_t \), the long rate with \( n (>1) \) periods to maturity is represented by \( R^n_t \) and \( c \) is a constant term premium. Equation (1) then states that the long rate, \( R^n_t \), is a weighted average of expected future short interest rates, \( r_t \), plus a constant, through time, term premium, \( c \).

It can be shown that the fundamental EH equation (1) may be transformed through subtraction of the short rate, \( r_t \), from both sides, into the following expression (setting the term premium \( c \) to zero, for simplicity):

\[
R^n_t - r_t = \sum_{i=1}^{n-1} \left( 1 - \frac{i}{n} \right) E_t \Delta r_{t,i}
\]

(2)

where \( S^n_t = R^n_t - r_t \) is the actual spread between the long and the short rate. Campbell and Shiller (1991) define the variable \( S^n_t = \sum_{i=1}^{n-1} (1 - i/n) \Delta r_{t,i} \) in expression (2) as the spread predicted by the model if agents have ‘perfect foresight’. The ‘perfect foresight spread’ (PFS) is thus constructed as a weighted average of changes in future short rates, using ex-post realisations of interest rates from the sample. Expression (2) then states that, according to the EH, the actual spread, \( S^n_t \), is the expected PFS, \( E_t S^n_t \). This formulation (2) is normally adopted when interest rates are believed to be non-stationary I(1) processes.
Adding to either of these expressions ((1),(2)) the rational market expectations (RE) assumption that the short rate is equal to its forecast plus a white-noise error term:

\[ r_t = E_t r_t + \eta_t \]  

(3)

has led to a number of tests of the EH proposed in the literature. One of the OLS regression tests suggested by Campbell and Shiller (1991) that we consider here is:

\[ S_t^n = \alpha + \beta S_t^n + \epsilon_t \]  

(4)

We expect \( \beta = 1 \) and \( \alpha = 0 \) under the null of the joint pure expectations hypothesis (EH with zero term premium \( c \)) with the RE assumption (thereby simply referred to as EH). The error term is serially correlated and follows an MA(\( n-2 \)) process and a GMM correction to the variance covariance matrix of the coefficients can be used.

We also consider the corresponding variance bound inequality, reformulated by Campbel and Shiller (1991) in terms of \( S_t^n \) and \( S_t^n \) as:

\[ Var(S_t^n) \geq Var(S_t^n) \]  

(5)

The variance (or standard deviation) of the interest rate spread may not exceed that of the PFS when the EH is true. Equivalently, the ratio of the corresponding standard deviations (or variances) must be less than 1. It is generally expected that the two tests (regression (4) and variance bound (5)) must give similar inferences.

2.2 The Campbell - Shiller VAR methodology

Campbell and Shiller (1991) show that the fundamental EH equation (1) may still be tested by replacing the RE assumption with an autoregressive scheme for predicting the perfect foresight spread (in expression (2)), as long as the actual spread \( S_t^n \) is included in the forecasting variables.
Assuming, first, that \(\Delta r_t\) and \(S_t^n\) are stationary stochastic processes, vector \(Z_t = \begin{bmatrix} S_t^n & \Delta r_t \end{bmatrix}^\top\) is also a stationary stochastic process. It is then assumed that a \(p\)-th order vector autoregression (VAR) representation of \(Z_t\) may be used to forecast the PFS. This bivariate VAR(\(p\)) may be written in companion form (as a 1st order VAR):

\[
z_{t+1} = Az_t + \omega_{t+1} \tag{6}
\]

where, \(A\), is a square \((2p \times 2p)\) matrix of coefficients and \(z_t\) is a \((2p \times 1)\) vector of regressors, \(z_t = \begin{bmatrix} \Delta r_t & \Delta r_{t-1} & \ldots & \Delta r_{t-p+1} & S_t^n & S_{t-1}^n & \ldots & S_{t-p+1}^n \end{bmatrix}^\top\). The companion form VAR may then be used to forecast (changes in) future short rates as required by the EH equation (2), by defining two \((2p \times 1)\) selection vectors \(e_1\) and \(e_2\) such that \(S_t^n = e_1'z_t\) and \(\Delta r_t = e_2'z_t\). These vectors choose the two key variables of interest, \(S_t^n\) and \(\Delta r_t\), through one element equal to 1, which is the 1st element, for \(e_1\), and the \(p+1\)st element (row) for \(e_2\), with all remaining \((2p-1)\) elements being zero.

Using then the VAR from equation (6) to forecast one period ahead:

\[
E_t\Delta r_{t+1} = e_2'E_tz_{t+1} = e_2'Az_t \tag{7}
\]

More generally, the ‘chain rule of forecasting’ for multi period expectations using the VAR implies:

\[
E_t\Delta r_{t+i} = e_2'Az_{t+i} \tag{8}
\]

The generalised VAR forecast for \(E_t\Delta r_{t+i}\) from equation (8) can then be replaced in the EH equation (2):

\[
S_t^n = E_tS_t^n = \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right)e_2'Az_i \tag{9}
\]
The forecast of the PFS using the VAR, in the right hand side of equation (9), is the ‘theoretical spread’, $S_t^\prime$. Therefore, if the EH (2) is true, the actual spread should be equal to the theoretical spread, $e_1'z_t = S_t^\prime$. From (9), Campbell and Shiller (1991) derive the following restrictions on the VAR parameters when the EH is true:

$$f(a) = e_1' - e_2'A \left[ I - \left( \frac{1}{n} \right)(I - A^n)(I - A)^i \right] (I - A)^i = 0 \quad (10)$$

The $(2p \times 1)$ non-linear restrictions, $f(a)$, expressed in (10) depend only on the estimated parameters, $\hat{A}$, of the unrestricted VAR model (6). A Wald test may then be used (as in Campbell and Shiller, 1987) to test their validity:

$$W = f(a) [\text{var}[f(a)]]^{-1} f(a)'$$

The variance of the restrictions, $\text{var}[f(a)]$, depends on the variance-covariance matrix of the VAR coefficients, $\Sigma$:

$$\text{var}[f(a)] = f_a(a)' \Sigma f_a(a) \quad (12)$$

where $f_a(a)$ is the first derivative of the restrictions evaluated at the coefficient estimates (from $A$).

In addition to the Wald test, the VAR methodology employs a small list of other measures of proximity between $S_t$ and $S_t^\prime$, that can be used to evaluate the EH, as it suggests through (9) that the two series must be equal. Besides the informal use of a graph of the two series, statistical measures include a ratio of standard deviations, the correlation statistic, and an OLS regression test:

$$sd(S_t^\prime)/sd(S_t^n) = 1 \quad (13)$$

$$\rho = \text{Corr}(S_t^\prime, S_t^n) = 1 \quad (14)$$
\[
S_i^n = \alpha + \beta S_i^n + \nu_i \tag{15}
\]

where we expect \( \alpha = 0 \) and \( \beta = 1 \).

### 3. Small sample bias and interest rate persistence

Considering again the previous test statistics, it can be shown that nearly all of them suffer from small sample bias that becomes quite severe in some cases, depending on the time series properties of interest rates and of course the size of the sample. All OLS based tests including the VAR regression tests potentially provide biased coefficient estimates (Bekaert et al., 1997). Similarly, the variance bounds family of tests (Flavin, 1983) also tends to distort inference in samples of typical size. This section focuses on the bias in the slope coefficient of the PFS regression test in (4). Briefly, in this case, the OLS coefficient estimates are biased due to the presence of a stochastic and not necessarily exogenous regressor (\( S_i^n \)).

First, we model the one period interest rate as a first order autoregressive (AR(1)) process. Besides the desirable property of greater analytical tractability (in the EH framework), the AR(1) model usually describes well interest rate data for actual economies. Thus, the one period interest rate, \( r_t \), is:

\[
r_t = \mu + \rho r_{t-1} + \varepsilon_t \tag{16}
\]

where \( |\rho| < 1 \) and the error term, \( \varepsilon_t \), is a normally and independently distributed process with mean zero and a constant variance, \( \sigma_{\varepsilon}^2 \).

When the EH (+RE) is true, the long rate and the spread are proportional to \( r_t \), hence, they too are persistent. In greater detail, when \( \mu = 0 \) (in 16) and the AR(1) is substituted out to forecast future one-period short rates in the fundamental EH equation (1) the long rate is:

\[
R^n = \left( \frac{1}{n} \right) r_t \sum_{j=0}^{n-1} \rho^j = \left( \frac{1}{n} \right) \left( \frac{1 - \rho^n}{1 - \rho} \right) r_t \tag{17}
\]
Accordingly, the spread, $S^a_t$, is also analogous to $r_t$:

\[ S^a_t = R^a_t - r_t = \left[ \frac{1}{n} \left( \frac{1 - \rho^a_t}{1 - \rho} \right) - 1 \right] r_t = \eta(\rho, n)r_t \tag{18} \]

where $\eta(\rho, n) = \left[ \frac{1}{n} \left( \frac{1 - \rho^a_t}{1 - \rho} \right) \right] - 1$.

Using the AR(1) representation for $r_t$, expression (18) implies that the spread is persistent and the persistence parameter is, again, $\rho$:

\[ S^a_t = \eta(\rho, n)(\rho r_{t-1} + \epsilon_t) = \rho \eta(\rho, n)r_{t-1} + \eta(\rho, n)\epsilon_t = \rho S^a_{t-1} + \nu_t \tag{19} \]

The PFS regression (4) and expression (19) essentially form the following system of equations:

\begin{align*}
S^a_t &= \alpha + \beta S^a_{t-1} + u_t \quad \text{where } u_t \sim \text{i.i.d}(0, \sigma_u^2) \tag{20} \\
S^a_t &= \xi + \rho S^a_{t-1} + v_t \quad \text{where } v_t \sim \text{i.i.d}(0, \sigma_v^2) \tag{21}
\end{align*}

The second equation implies that $S^a_t$ is not fixed in repeated samples, violating an OLS requirement for the estimation of the PFS equation (20). Moreover, the error terms, $u_t$ and $v_t$, are contemporaneously correlated with covariance $\sigma_{uv}$.

Standard econometric theory tells us that if $S^a_t$, the independent variable in the PFS regression, is stochastic then this is not necessarily problematic for the typical properties of the OLS estimator. The Gauss-Markov theorem may still hold provided the error process of the regressor is independent of the error process of the regression itself. However, even in that case, the small sample properties of the OLS estimator may be poor.$^1$

Nevertheless, in this system, there is also a correlation ($\sigma_{uv}$) between the disturbances of the independent variable ($S^a_t$) and the disturbances of the PFS regression (20) itself and therefore the Gauss-Markov theorem no longer applies. In the particular example, this is because the actual spread inherits not only the persistence but also the stochastic component of the short rate, $\epsilon_t$. 

---

$^1$
from the AR(1) model (16). The PFS on the left hand side of (20) also contains that component as it includes $\Delta r_{t+1}$ (from EH (2)). Therefore, the OLS coefficient estimates, $\hat{\alpha}$ and $\hat{\beta}$, are consistent but biased, in samples of typical size. Moreover the sampling distribution of this small-sample estimator is not standard.

This problem is well known in the finance literature, as it applies to forecasting equations often used in tests of asset pricing theories, including the PFS regression (e.g., Stambaugh 1986, 1999; Mankiw and Shapiro, 1986). Stambaugh (1986) has shown that if the stochastic regressor is also persistent (as $S^r$ is here) then the bias in estimating the OLS slope coefficient depends on the bias in estimating the persistence coefficient, $\rho$, as follows:

\[
\text{E}(\hat{\beta} - \beta) = \left( \frac{\sigma_{uv}}{\sigma^2_u} \right) \text{E}(\hat{\rho} - \rho) \tag{22}
\]

The bias in estimating the autocorrelation coefficient ($\rho$) of an autoregressive process is negative (downward) (Marriott and Pope, 1954). The closer the true coefficient, $\rho$, is to 1, the more severe this type of bias becomes. Moreover, in the PFS regression, the covariance, $\sigma_{uv}$, is negative so the slope bias (in (22)) is positive (upward).

Bekaert et al. (1997), illustrate that in the case of the PFS regression the slope bias expression (22) will implicitly involve estimation of higher order autocorrelation coefficients (due to the future short rates involved in constructing the PFS) as well. They use Kendall’s (1954) analytical approximations for the bias in higher order autocorrelation coefficients to reach a more specific expression (than 22) of the bias in $\hat{\beta}$, in the PFS regression. The bias is expressed as:

\[
\text{E}(\hat{\beta} - 1) = \frac{1}{n \eta(\rho, n)} \sum_{j=1}^{\infty} \theta_j = \frac{-1}{n(1 - \rho)} \sum_{j=1}^{\infty} \theta_j \geq 0 \tag{23}
\]

where $\eta(\rho, n) = \left( \frac{1}{n} \right) \left( \frac{1 - \rho^n}{1 - \rho} \right)^{-1}$ and $\theta_j$ is Kendall’s approximation to the estimated autocorrelation coefficient bias:

1 E.g. see pp. 35-36 in Cuthbertson et al. (1992).
\[ \theta_j = E(\hat{\rho}_j - \rho_j) = - \left( \frac{1}{T-(n-1)} \right) \left[ \frac{(1+\rho)(1-\rho')}{(1-\rho^2)} \right] + 2j\rho' \]  

(24)

where \( j \) is the order of autocorrelation. Hence, in this case: \( E(\hat{\beta} - 1) = f(n,T,\rho) \). The estimated \( \hat{\beta} \) is positively biased, above 1. The bias increases with higher \( \rho \), smaller \( T \) and smaller \( n \). The bias in expression (23) is particularly strong for very high autocorrelation coefficients, \( \rho \), taking values between 0.90 and 0.99. Bekaert et al. (1997) show that the empirical distribution is highly dispersed and skewed to the right. It takes almost as many as 2,000 observations for symmetry.

Considering, now, the VAR tests and statistics, Bekaert et al. (1997) show that the VAR OLS coefficients are all biased, depending on persistence. The bias passes through to nearly all the VAR-based tests and metrics. In summary, they show that: a) to a first order approximation, \( \text{Corr}(S'_t, S_t) = 1 \), the coefficient of correlation between the spread \( S_t \) and the ‘theoretical spread’ \( S'_t \) is unbiased for small samples and b) the standard deviation ratio \( sd(S'_t)/sd(S_t) \) is biased (upwards) as \( \rho \) rises, exceeding its theoretical value of \( 1^2 \). Bekaert and Hodrick (2001) show that the small sample distribution of the Wald test of the EH restrictions for the VAR coefficients is severely distorted. The Wald statistic tends to overreject the EH (biased upwards).

4. Monte Carlo experiments
4.1 Monte Carlo Structure

The results discussed in the previous section show that small sample inference is essential for testing the EH, especially when interest rate data display a high degree of persistence. Given that asymptotic inference is misleading in finite samples, well-designed Monte Carlo simulation (MCS) experiments can be used to correct for small sample bias (e.g., Longstaff, 2000).

Analytical values from closed form estimates can either be tedious to reach (in the case of the VAR) or of doubtful validity for high \( \rho \). Moreover, estimates of the bias in the coefficients are not sufficient for statistical inference given also the skewness of the small sample distributions.

\[ \begin{array}{c}
\text{Campbel and Shiller (1991) also report Monte Carlo evidence for the small sample distributions of the VAR statistics. They find a smaller bias, not large enough to overturn the rejection of the EH for their data.}
\end{array} \]
We use Monte Carlo simulation experiments to explore the sensitivity of the small sample distributions of the different EH tests, described in section 2, to changing persistence levels. The AR(1) process for \( r_t \) in (16) is central to the Monte Carlo analysis, as we are theoretically interested in the responsiveness of the shape of these distributions to persistence, \( \rho \). However, researchers wishing to test actual data might bootstrap their estimates and correct the data generating process for the bias in \( \rho \), while also adopting a richer representation of reality with a VAR or VAR-GARCH model for the short rate (e.g., Longstaf, 2000; Bekaert et al., 1997).

We generate data under the null of EH (=EH+RE). This requires us to use the fundamental EH equation (1) to generate the long rate \( R^n_t \) (for \( n=12 \)), which in turn requires us to substitute expected future one period rates. We, first, generate \((T+200)\) observations of the one period rate, \( r_t \), using the AR(1) model in equation (16) with normally distributed random errors and then recursively substitute for future short rates (RE) in the long rate model (EH). For the generated series, we then discard the first 200 observations to minimise the influence of initial conditions. The actual spread \( S^{12}_t \) is then the difference between the two interest rate series, \( R^{12}_t \) and \( r_n \), and we construct the PFS \( S^{12}_t \) (from the EH equation (2)) as the weighted average of changes in the forwarded short rates, substituting recursively the AR(1) model. Having generated series for the PFS and the actual spread, complying with the EH, we estimate in each experiment: 1) the PFS OLS regression test (4) and 2) in a new round of Monte Carlo experiments, the standard deviation ratio of the two series (5). In a third round of experiments, with the same process for data generation, and after calculating the actual spread \( S^{12}_t \) and \( \Delta r_t \), we estimate a VAR(1) model and use the coefficient estimates to construct the theoretical spread, \( S'_t \). We then estimate in each replication the corresponding OLS regression test (15) and the VAR metrics (correlation and standard deviation ratio). For all these experiments, we derive the small sample distributions of the coefficients and statistics, replicating the exercise 5,000 times. We repeat the 5,000 Monte Carlo replications for artificial samples of \((T=)\) 124, 524, 2,000 and 20,000 observations. ‘Antithetic variates’ are used in all experiments with the exception of the variance bound test for which we repeat 10,000 times (instead of 5,000) the Monte Carlo trials. For the AR(1) parameters in (16) we assume fixed values for \( \mu = 1.24 \) and \( \sigma^2_e = 2.28 \), with \( \rho \) changing. We also validate the code by trying the same set of parameter values for the AR(1) process as those used in Bekaert et al. (1997) in their similar experiment. For the slope of the PFS regression test, we find beta=1.47.
against their 1.51 with a standard deviation $s.d.=0.62$ against 0.65 (further similarities are documented in figure 1 and figure 2).
4.2 Simulation results

4.2.1 The perfect foresight spread OLS test

Table 1 reports Monte Carlo results (for 5,000 trials) that outline the small sample distribution of the slope coefficient $\beta$ in the PFS regression test from experiments in which (AR(1)) interest rates (from equation (16)) were generated, as described in the previous section, for different $\rho$. We also explore larger sample properties of the test for samples of 524, 2,000 and 20,000 observations.

Table 1. OLS slope coefficient $\beta$ for PFS 12: $\rho$ and sample size sensitivity

Panel A, reports Monte Carlo results for different samples and for a broad range of $\rho$ between 0 and 1. For each $\rho$, the mean slope coefficient $\hat{\beta}$ and the standard deviation around this estimate are reported. The first value reported is from OLS estimation that includes a constant term, while the value underneath denotes OLS estimation without a constant term. Panel B, reports results for very high values of $\rho$ and a sample of 124 observations, for three quartile levels of the small sample distribution (besides the mean and the standard deviation).

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\beta$ (s.e.)</th>
<th>$\beta$ (s.e.)</th>
<th>$\beta$ (s.e.)</th>
<th>$\beta$ (s.e.)</th>
<th>$\beta$ (s.e.)</th>
<th>$\beta$ (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.919 (0.044)</td>
<td>0.981 (0.015)</td>
<td>0.995 (0.007)</td>
<td>1.00 (0.002)</td>
<td>0.991 (0.044)</td>
<td>0.979 (0.016)</td>
</tr>
<tr>
<td>0.25</td>
<td>0.924 (0.053)</td>
<td>0.982 (0.023)</td>
<td>0.995 (0.011)</td>
<td>1.00 (0.004)</td>
<td>0.912 (0.056)</td>
<td>0.979 (0.023)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.935 (0.078)</td>
<td>0.986 (0.038)</td>
<td>0.997 (0.019)</td>
<td>1.00 (0.006)</td>
<td>0.914 (0.081)</td>
<td>0.980 (0.038)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.979 (0.146)</td>
<td>0.998 (0.077)</td>
<td>0.998 (0.041)</td>
<td>1.00 (0.013)</td>
<td>0.927 (0.150)</td>
<td>0.982 (0.078)</td>
</tr>
<tr>
<td>1</td>
<td>0.911 (0.159)</td>
<td>0.979 (0.079)</td>
<td>0.995 (0.041)</td>
<td>1.00 (0.013)</td>
<td>0.911 (0.159)</td>
<td>0.979 (0.080)</td>
</tr>
</tbody>
</table>

*Note: 1st row for each value of the autocorrelation coefficient, $\rho$, refers to standard OLS value of the slope and the second row refers to the restricted-constant OLS value of the slope.
Overall, the results are in accordance with the analytic approximation\(^3\) and Monte Carlo evidence in Bekaert et al. (1997). For very high \(\rho\), above 0.85, (in Panel B) we find significant positive bias and distortions of the small sample distribution of the estimator. However, for lower degrees of persistence (Panel A), a negative bias is also present. For example, the analytical approximation estimate for \(\rho = 0.75\) and a sample of 124 observations is \(\hat{\beta} = 1.08\), which implies a very small positive bias (against the theoretical value of 1). The mean estimate in our Monte Carlo results is 0.98, suggesting a very small negative bias in the estimate of the slope coefficient, \(\beta\). The bias is present for every \(\rho\) in Panel A of the table, for samples of 124 and 524 observations. It is due to small sample influences, other than that from the persistence of interest rates, perhaps also to be expected.

The results in Panel A of table 1 suggest two patterns as we increase \(\rho\) and the sample size, \(T\). As \(\rho\) varies from 0 to 0.75 the mean slope, \(\hat{\beta}\), (and the standard deviation around it) rises from 0.92 to 0.98 for 124 observations and similarly for samples up to 2000 observations\(^4\). Thus as autocorrelation rises, the implied positive bias in estimating \(\beta\) (described in Bekaert et al., 1997) rises, partly offsetting the negative bias found in small samples. Considering now larger samples, the value of the slope \(\beta\) approximates the theoretical value of 1, confirming the consistency property of the OLS estimator. It is striking that even for a substantially large sample of 2000 observations the mean slope, which is very close to 1, still slightly rises with \(\rho\), at the third decimal point. The slope distribution collapses exactly to 1, with a very small standard deviation, only for a sample of 20,000 observations.

Panel B of table 1, reports results for high persistence, with \(\rho\) between 0.90 and 0.99 and a sample of 124 observations. The positive bias now becomes large (e.g. mean \(\hat{\beta} = 1.63\) for \(\rho = 0.96\)) and the small sample distribution is widely dispersed (standard deviation = 0.69). We find that 97.5% of the slope coefficient estimates for \(\rho = 0.96\) are below 3.11, implying that if a slope coefficient is estimated to be as high as 3 in actual data for a market of the same degree of interest rate persistence, the PEH should still not be rejected, using this table. The results suggest that, using small sample, rather than standard asymptotic, inference, it is more difficult to reject the null

---

\(^3\) This is proposition 3 in Bekaert et al. (1997), stated in equation (23) here.

\(^4\) We also report the unit root case, for \(\rho = 1\), which behaves differently. The analytical approximation of the bias refers only to values smaller than 1.
hypothesis when interest rate persistence is high due to the large dispersion of the small sample distribution\(^5\).

Figure 1. PFS test slope coefficient: small sample against asymptotic distribution

![Density curves (scaled) for small sample vs Asymptotic Normal distribution](image)

Note: MCS results for sample size \(T=124\) observations and \(\rho=0.928\). NW-s.e. =0.092.

Figure 2. Monte Carlo density curves for larger samples

![Monte Carlo density curves for larger samples](image)

We repeated all the above experiments *excluding* this time the constant term from the OLS regressions. We find that for high \(\rho\), the mean estimate of \(\beta\), in the small sample distribution (in table 1) and the median (which is not reported in the table) are closer to 1 (see figure 3) than they

were in standard OLS. The bias is substantially reduced though not eliminated. However, the standard deviation of the small sample distribution is again high. Therefore, for high values of $\rho$, the slope coefficient bias is almost eliminated when the constant term is excluded from OLS regressions but it is still more difficult to reject the null hypothesis (EH), than it is in large samples. The distinction between the two estimations is not important for lower $\rho$ between 0 and 0.75 (panel A).

Figure 3. Shift in the mean of the small sample distribution: zero OLS constant term

![Graph showing the shift in the mean of the small sample distribution](image)

Table 2 reports Monte Carlo evidence for the sensitivity of the small sample distribution across different spread horizons$^6$. The spread is calculated between the (EH-constructed) long rate for $n = 12, 9, 6, 3, 2$ and the AR(1) one period rate. The sample consists of 124 observations and $\rho$ is fixed to 0.93 ($\mu = 1.24$).

$^6$ For a study of market efficiency tests across different horizons see Campbell (2001).
Table 2. OLS test for PFS across different spreads: $\rho=0.93, T=124$.

<table>
<thead>
<tr>
<th>Spread maturity, $n$</th>
<th>Mean $\hat{\beta}$ (s.e.) (analytic)</th>
<th>t - ratio against 1</th>
<th>Quartiles: median</th>
<th>0.975</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1.265 (0.416) (1.415)</td>
<td>0.637</td>
<td>1.241</td>
<td>2.119</td>
<td>0.619</td>
<td>0.528</td>
<td>0.420</td>
</tr>
<tr>
<td>9</td>
<td>1.317 (0.464) (1.429)</td>
<td>0.684</td>
<td>1.277</td>
<td>2.314</td>
<td>0.626</td>
<td>0.534</td>
<td>0.434</td>
</tr>
<tr>
<td>6</td>
<td>1.375 (0.517) (1.443)</td>
<td>0.725</td>
<td>1.319</td>
<td>2.534</td>
<td>0.631</td>
<td>0.537</td>
<td>0.436</td>
</tr>
<tr>
<td>3</td>
<td>1.436 (0.589) (1.459)</td>
<td>0.741</td>
<td>1.348</td>
<td>2.831</td>
<td>0.634</td>
<td>0.542</td>
<td>0.442</td>
</tr>
<tr>
<td>2</td>
<td>1.448 (0.628) (1.465)</td>
<td>0.714</td>
<td>1.344</td>
<td>2.964</td>
<td>0.622</td>
<td>0.525</td>
<td>0.418</td>
</tr>
</tbody>
</table>

Note: ‘Analytical’ refers to the theoretical approximation, using proposition 3 in Bekaert et al. (1997).

Results in table 2 show that as the maturity ($n$) of the long rate in the spread $S^n$ becomes shorter, the mean $\hat{\beta}$ (and the median) from the MCS experiments and the standard deviation tend to rise. In the case of $S^2$, the mean Monte Carlo estimate of $\hat{\beta}$ is 1.45 (compared to the analytical estimate of 1.47), against 1.27 found for $S^{12}$. These results are broadly in agreement with the analytical approximations: as $n$ rises $E(\hat{\beta} - 1)$ (in equation (23)) becomes smaller. Therefore, in all the previous Monte Carlo results in table 1, for which $S^{12}$ was used as a benchmark case, the bias in $\hat{\beta}$ is understated when one refers to other spreads.
4.2.2 Variance bounds

In the next round of Monte Carlo experiments we explore the small sample distribution of the variance bound ratio \((\frac{s.d.(\bar{S}_t)}{s.d.(\tilde{S}_t^n)})\), for the same set of values for changing \(\rho\) and \(T\). We use again the AR(1) one period rate to generate the long rate under the null hypothesis (EH), for 10,000 replications\(^7\).

Table 3: Volatility ratio for changing \(\rho\) and \(T\)

The ratio \(VR = \frac{s.d.(\bar{S}_t^n)}{s.d.(\tilde{S}_t^n)}\) is calculated over 10,000 trials, for \(n=12\). Panel A, covers a wide range of values for \(\rho\), between 0 and 1 for samples of 124, 524, 2,000 and 20,000 observations. Panel B considers only high \(\rho\) for a small sample of 124 observations.

<table>
<thead>
<tr>
<th>(\rho)</th>
<th>Panel A</th>
<th>(n)</th>
<th>Mean VR (s.e.)</th>
<th>% of ratios &gt;1 (out of all trials)</th>
<th>Mean VR (s.e.)</th>
<th>Mean VR (s.e.)</th>
<th>Mean VR (s.e.)</th>
<th>Panel B</th>
<th>(n)</th>
<th>Mean VR (s.e.)</th>
<th>Quartile 0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0)</td>
<td>Larger samples: Number of Observations ((T)) = 124</td>
<td>124</td>
<td>1.000 (0.025)</td>
<td>45.8%</td>
<td>0.967 (0.008)</td>
<td>0.960 (0.004)</td>
<td>0.958 (0.001)</td>
<td>(0.85)</td>
<td>0.662 (0.077)</td>
<td>0.842</td>
<td>0.535</td>
</tr>
<tr>
<td>(0.25)</td>
<td></td>
<td>524</td>
<td>0.971 (0.029)</td>
<td>15.7%</td>
<td>0.940 (0.012)</td>
<td>0.934 (0.006)</td>
<td>0.932 (0.002)</td>
<td>(0.93)</td>
<td>0.473 (0.085)</td>
<td>0.683</td>
<td>0.338</td>
</tr>
<tr>
<td>(0.50)</td>
<td></td>
<td>2,000</td>
<td>0.918 (0.040)</td>
<td>3.4%</td>
<td>0.892 (0.018)</td>
<td>0.886 (0.009)</td>
<td>0.884 (0.003)</td>
<td>(0.96)</td>
<td>0.336 (0.078)</td>
<td>0.522</td>
<td>0.218</td>
</tr>
<tr>
<td>(0.75)</td>
<td></td>
<td>20,000</td>
<td>0.789 (0.063)</td>
<td>0.6%</td>
<td>0.772 (0.030)</td>
<td>0.770 (0.016)</td>
<td>0.769 (0.005)</td>
<td>(0.99)</td>
<td>0.335 (0.077)</td>
<td>0.522</td>
<td>0.218</td>
</tr>
<tr>
<td>(1)</td>
<td>High Persistence Number of Observations ((T)) = 124</td>
<td></td>
<td>0.142 (0.018)</td>
<td>0.0%</td>
<td>0.069 (0.005)</td>
<td>0.036 (0.001)</td>
<td>0.011 (0.000)</td>
<td>(1)</td>
<td>0.142 (0.018)</td>
<td>0.180</td>
<td>0.104</td>
</tr>
</tbody>
</table>

\(^7\) We double the number of replications to 10,000, in this case, as the ‘antithetics’ technique does not make a difference for calculating standard deviations.
We find that the mean volatility ratio estimated over all trials, for a sample of 124 observations and for $\rho = 0$, is exactly equal to 1 and the 97.5% quartile (which is not reported in the table) is at 1.06, which violates the bound. This seems surprising at first, given that the theoretical value of 1 is the exact upper bound and that data were generated under the null. The violation of the bound is observed for higher $\rho$ as well. This is the type of violation of the bound explained in J. Flavin’s known paper (1983), though for interest rate, rather than spread, volatility ratios. In small samples the ratio is biased upwards tending to violate the bound and reject the EH ‘too often’, when the null hypothesis (EH) is true.

However, in our Monte Carlo results, as $\rho$ rises, the mean ratio tends to fall. For $\rho = 0.75$ and a sample of 124 observations, the mean ratio is equal to 0.79 against 0.91 for $\rho = 0.5$ and correspondingly the 97.5% quartile level is found at 0.94 against 1.01 (not in the table). The standard error around the mean Monte Carlo estimate also rises with $\rho$. Therefore, in opposite direction to the upward small sample bias there is also a downward bias to the estimated ratio, which is stronger when persistence rises. The ‘persistence effect’ is more evident in panel B (of table 3) for $\rho > 0.85$.

The second column of the table (3) shows the ‘opposing’ biases more clearly. When $\rho$ is low there are frequent violations of the bound (e.g., at $\rho = 0$, 46% of the ratios are above 1). As persistence rises, however, it is found that violations fall towards 0, as a result of the dominating, downward, ‘persistence bias’ found in these experiments.

For larger samples, biases tend to be eliminated. However, the bias due to higher persistence still exists given that the ratio tends to fall when persistence is higher. Even for a very large sample of 20,000 observations, at $\rho = 0.50$, the mean ratio is 0.93, falling to 0.77 at $\rho = 0.75$. Overall, large sample results do not collapse to a specific value, unlike the case of OLS, where the estimator was found consistent. Unlike the OLS case, the variance bounds ratio is not expected theoretically to converge to a specific value and it must only satisfy the inequality.

---

8 The unit root scenario, $\rho = 1$, is again a special case and the mean ratio dramatically falls to 0.14.
Table 4. Volatility ratios across different spreads: \( \rho = 0.93, T=124 \)

<table>
<thead>
<tr>
<th>Spread maturity, ( n )</th>
<th>Mean VR (s.e.)</th>
<th>median</th>
<th>0.975</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.473 (0.085)</td>
<td>0.457</td>
<td>0.683</td>
<td>0.362</td>
<td>0.351</td>
<td>0.338</td>
</tr>
<tr>
<td>9</td>
<td>0.409 (0.076)</td>
<td>0.397</td>
<td>0.593</td>
<td>0.308</td>
<td>0.295</td>
<td>0.284</td>
</tr>
<tr>
<td>6</td>
<td>0.331 (0.063)</td>
<td>0.322</td>
<td>0.478</td>
<td>0.243</td>
<td>0.232</td>
<td>0.221</td>
</tr>
<tr>
<td>3</td>
<td>0.224 (0.046)</td>
<td>0.218</td>
<td>0.331</td>
<td>0.159</td>
<td>0.151</td>
<td>0.142</td>
</tr>
<tr>
<td>2</td>
<td>0.167 (0.035)</td>
<td>0.163</td>
<td>0.245</td>
<td>0.117</td>
<td>0.110</td>
<td>0.105</td>
</tr>
</tbody>
</table>

Table 4, reports results for different maturity horizons, \( n \), of the spreads \( S^n \). The level of persistence is fixed at \( \rho = 0.93 \) in all experiments, for samples of 124 observations. Similarly to our previous results for the OLS regression test, when the long rate maturity, \( n \), becomes shorter, the bias is larger and the mean ratio for \( S^2 \) falls to 0.17 against 0.47 for \( S^{12} \). Again, the results reported in table 3 and referring to \( S^{12} \), understate the ‘persistence bias’ for shorter maturity spreads.
4.2.3 VAR tests and metrics

In this round of Monte Carlo experiments we explore the properties of the VAR model. Using the same parameter values, we generate $r_t$ as an AR(1) process and the long rate according to the EH, as before. In each trial, we generate data under the null, we estimate by OLS a bivariate VAR(1) model:

$$S^n_{t+1} = a_{11}S^n_t + a_{12}\Delta r_t + u_{1,t+1}$$

$$\Delta r_{t+1} = a_{21}S^n_t + a_{22}\Delta r_t + u_{2,t+1}$$

We then calculate the ‘theoretical spread’ and the associated statistics of proximity to the actual spread: i) the standard deviation ratio, $sd(S^\prime)/sd(S)$, ii) the correlation statistic, $Corr(S^\prime,S)$, iii) the slope coefficient from the OLS regression of $S^\prime$ against $S$, and iv) the Wald test of restrictions for the VAR coefficients, $A$, implied by the EH (in equation (10)) and their (Newey-West) robust variance covariance matrix.

Table (5) reports results describing the small-sample distribution of $sd(S^\prime)/sd(S)$ and its sensitivity to changing values of $\rho$ and sample size, $T$. In accordance with the analytic approximation in Bekaert et al. (1997), the ratio $sd(S^\prime)/sd(S)$ is biased upwards in a small sample, when $\rho$ is rising. For $\rho$ up to 0.75, in panel A, the mean ratio, $sd(S^\prime)/sd(S)$, is only slightly above the theoretical value of 1 and rising with $\rho$. As the sample size, $T$, rises, for each $\rho$, the mean ratio falls closer to 1.

The bias is more pronounced for $\rho$ between 0.85 and 0.96, in panel B, as the mean standard deviation ratio rises from 1.01 to 1.65. The small sample distribution becomes more dispersed, with standard deviations around the mean ratios correspondingly rising from 0.18 to 0.66. For $\rho = 0.96$, theoretical standards of ‘closeness’ to 1 would say that $sd(S^\prime)/sd(S) = 1.6$, is not ‘close enough’, however, under small-sample inference using table 5, 1.6 is the mean ratio, (with a high standard deviation of 0.66) and the EH is correctly supported.
Table 5: VAR metrics: $sd(S_t')/sd(S_t)$ for PFS 12: $\rho$ and $T$ sensitivity

<table>
<thead>
<tr>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Larger samples:</strong></td>
<td><strong>High Persistence</strong></td>
</tr>
<tr>
<td>Number of Observations ($T$)</td>
<td>Number of Observations ($T$)</td>
</tr>
<tr>
<td>$124$</td>
<td>$124$</td>
</tr>
<tr>
<td>$524$</td>
<td></td>
</tr>
<tr>
<td>$2,000$</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>$\rho$</td>
</tr>
<tr>
<td>Mean ratio</td>
<td>Mean ratio</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(s.e.)</td>
</tr>
<tr>
<td>$0$</td>
<td>$1.001$ (0.011)</td>
</tr>
<tr>
<td>$0.25$</td>
<td>$1.002$ (0.018)</td>
</tr>
<tr>
<td>$0.50$</td>
<td>$1.004$ (0.035)</td>
</tr>
<tr>
<td>$0.75$</td>
<td>$1.027$ (0.094)</td>
</tr>
</tbody>
</table>

Table 6: VAR metrics: $sd(S_t')/sd(S_t)$ across spreads, $\rho = 0.93$, $T=124$

<table>
<thead>
<tr>
<th>Spread maturity, $n$</th>
<th>Mean ratio (s.e.)</th>
<th>Median</th>
<th>0.975</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1.265 (0.353)</td>
<td>1.253</td>
<td>1.979</td>
<td>0.701</td>
<td>0.626</td>
<td>0.546</td>
</tr>
<tr>
<td>9</td>
<td>1.347 (0.447)</td>
<td>1.317</td>
<td>2.292</td>
<td>0.667</td>
<td>0.580</td>
<td>0.484</td>
</tr>
<tr>
<td>6</td>
<td>1.416 (0.528)</td>
<td>1.347</td>
<td>2.604</td>
<td>0.659</td>
<td>0.572</td>
<td>0.469</td>
</tr>
<tr>
<td>3</td>
<td>1.523 (0.618)</td>
<td>1.418</td>
<td>2.970</td>
<td>0.700</td>
<td>0.611</td>
<td>0.515</td>
</tr>
<tr>
<td>2</td>
<td>1.581 (0.6513)</td>
<td>1.487</td>
<td>3.090</td>
<td>0.716</td>
<td>0.637</td>
<td>0.525</td>
</tr>
</tbody>
</table>
Table 6 presents results for the standard deviation ratio across different maturity horizons. The mean ratio (and standard deviation around it) rises as the long rate maturity, \( n \), of the spread, \( S' \), shortens. This result is not in accordance with the corresponding analytic approximation in Bekaert et al. (1997), according to which this statistic does not (to a first approximation) depend on the maturity horizon, \( n \). However, a similar pattern can be observed in their Monte Carlo results.

In the same Monte Carlo exercise we estimate the correlation coefficient, \( \text{Corr}(S'_t, S_t) \), and find that it is unbiased with respect to rising persistence. As expected from theory, it collapses almost perfectly to its theoretical value (=1), for all \( \rho \). Moreover, in contrast with theory, we find that the mean \( \text{Corr}(S'_t, S_t) \) varies to a small degree across maturity horizons, \( n \). We do not report these results in a table, as any variations are too small.

Looking now at the main statistical test of the VAR methodology, the Wald test of the VAR cross-equation restrictions, table 7 shows results from the same experiments (for varying \( \rho \) and sample sizes, \( T \)). Given that there are no tractable analytic approximations for the theoretical bias in this statistic, with respect to persistence, we only rely on results from Monte Carlo evidence. Similarly to Bekaert and Hodrick (2001), we find that the small sample properties of the Wald test of the EH restrictions are very poor.
<table>
<thead>
<tr>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Larger samples:</strong></td>
<td><strong>High Persistence</strong></td>
</tr>
<tr>
<td>Number of Observations (T) =</td>
<td>T=124</td>
</tr>
<tr>
<td>124</td>
<td>524</td>
</tr>
<tr>
<td>ρ</td>
<td>Mean (s.e.)</td>
</tr>
<tr>
<td>quartile (1-right tail)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>11.562</td>
</tr>
<tr>
<td>0.25</td>
<td>11.799</td>
</tr>
<tr>
<td>0.50</td>
<td>12.670</td>
</tr>
<tr>
<td>0.75</td>
<td>15.145</td>
</tr>
</tbody>
</table>
Table 8. VAR tests: Wald test across spreads, $\rho =0.93$, $T=124$

The first row of the table quotes critical values from the $\chi^2(2)$ distribution, from standard statistical tables.

<table>
<thead>
<tr>
<th>Spread</th>
<th>Mean (s.e.)</th>
<th>quartiles</th>
<th>0.95</th>
<th>0.99</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2(2$ d.f.)</td>
<td>Na</td>
<td>1.386</td>
<td>5.990</td>
<td>9.210</td>
<td>0.051</td>
<td>0.020</td>
</tr>
<tr>
<td>$S^2$</td>
<td>3.162 (3.470)</td>
<td>2.103</td>
<td>9.672</td>
<td>16.455</td>
<td>0.081</td>
<td>0.031</td>
</tr>
<tr>
<td>$S^l$</td>
<td>3.412 (3.992)</td>
<td>2.129</td>
<td>10.774</td>
<td>19.025</td>
<td>0.075</td>
<td>0.031</td>
</tr>
<tr>
<td>$S^0$</td>
<td>3.776 (5.091)</td>
<td>2.170</td>
<td>12.409</td>
<td>23.812</td>
<td>0.081</td>
<td>0.029</td>
</tr>
<tr>
<td>$S^r$</td>
<td>4.475 (6.924)</td>
<td>2.259</td>
<td>15.858</td>
<td>33.398</td>
<td>0.086</td>
<td>0.033</td>
</tr>
<tr>
<td>$S^{12}$</td>
<td>5.189 (11.521)</td>
<td>2.247</td>
<td>19.851</td>
<td>41.331</td>
<td>0.062</td>
<td>0.020</td>
</tr>
</tbody>
</table>

Overall, we find a pronounced upward bias in the mean Wald test statistic, estimated over 5,000 Monte Carlo trials, and a more dispersed distribution, as $\rho$ rises, even in larger samples. For fixed $\rho$, the bias falls, as $T$ rises. More particularly, when persistence, $\rho$, is low and samples are large the Wald test statistic is still close to the critical values suggested by the $\chi^2$ distribution, in panel A (table 7). For example, for large $T=2,000$, when $\rho =0.5$, the 95% quartile level is at 6.21 against the $\chi^2$ critical value of 5.99. However, when $\rho$ is raised to 0.75, for the same $T$, the 95% quartile level rises to 14.65.

Panel B (in table 7) shows that for very high values of $\rho$ and for a sample of 124 observations, the small sample distribution of the Wald statistic is considerably distorted. Under small sample inference and for $\rho = 0.96$, the EH should only be rejected for Wald test values higher than 22.12 (against the standard inference value of 5.99), at the 95% quartile. Hence using the Wald test under asymptotic inference the EH tends to be over-rejected in small samples.

Next, we explore in Monte Carlo results in table 8, differences across changing maturity horizons, $n$, for the spread, fixing $\rho = 0.93$, for a sample of 124 observations. The bias in the mean Wald test statistic is higher for longer horizon spreads and the 95% quartile level rises to 19.85 for $S^{12}$ against only 9.67, for $S^2$. Therefore, unlike all previous tests, $S^{12}$ which was used in all the
experiments reported in table 7, overstates the degree of ‘persistence bias’ that would apply to shorter maturity spreads.

We examine last the OLS regression test of the ‘theoretical spread’ $S_\tau$ against $S_\tau$ in Monte Carlo results reported in table 9. It should be noted that during these experiments, OLS estimation failed to a great extent in large samples. Therefore, our results focus on a small sample of 124 observations, for different $\rho$.

Table 9. OLS $\hat{\beta}$ from regression of $S_\tau$ against $S_\tau$, $T=124$, high $\rho$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Mean $\beta$ (s.e.)</th>
<th>t-ratio</th>
<th>quartiles: median</th>
<th>0.975 (1-right tail)</th>
<th>0.05 (left tail)</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000 (0.011)</td>
<td>0.027</td>
<td>1.002</td>
<td>1.018</td>
<td>0.980</td>
<td>0.975</td>
<td>0.969</td>
</tr>
<tr>
<td>0.50</td>
<td>1.004 (0.036)</td>
<td>0.104</td>
<td>1.007</td>
<td>1.062</td>
<td>0.941</td>
<td>0.926</td>
<td>0.902</td>
</tr>
<tr>
<td>0.75</td>
<td>1.025 (0.095)</td>
<td>0.263</td>
<td>1.032</td>
<td>1.190</td>
<td>0.856</td>
<td>0.817</td>
<td>0.777</td>
</tr>
<tr>
<td>0.85</td>
<td>1.076 (0.178)</td>
<td>0.430</td>
<td>1.081</td>
<td>1.405</td>
<td>0.774</td>
<td>0.724</td>
<td>0.653</td>
</tr>
<tr>
<td>0.93</td>
<td>1.263 (0.354)</td>
<td>0.745</td>
<td>1.252</td>
<td>1.977</td>
<td>0.699</td>
<td>0.619</td>
<td>0.540</td>
</tr>
<tr>
<td>0.96</td>
<td>1.642 (0.665)</td>
<td>0.924</td>
<td>1.591</td>
<td>3.062</td>
<td>0.654</td>
<td>0.516</td>
<td>0.331</td>
</tr>
<tr>
<td>0.99</td>
<td>4.068 (2.521)</td>
<td>1.217</td>
<td>3.679</td>
<td>9.810</td>
<td>0.568</td>
<td>0.159</td>
<td>-0.477</td>
</tr>
</tbody>
</table>

The results show that a positive bias in the slope coefficient is present and the small sample distribution is highly dispersed. The mean (and the median) $\hat{\beta}$ rises above 1 as $\rho$, rises towards 0.75. The standard deviation around $\hat{\beta}$ also rises.

Focusing on high $\rho$, table 9 shows a more pronounced slope coefficient bias that is even larger than that recorded for $\hat{\beta}$ in the PFS test, in table 1. Table 9 shows that, for a real market in which $\rho$ is near 0.92, if we find $\hat{\beta} = 1.90$ then the EH must still not be rejected, using small sample
inference, at the 97.5% quartile level. Moreover, it is interesting to note that in larger samples the mean $\hat{\beta}$ found here converges to 1 more quickly than it did in the PFS regressions.

**Figure 5. Monte Carlo results: OLS $\hat{\beta}$ ($S_t'$ against $S_t$) histograms for high and low $\rho$**

![Monte Carlo histograms](image)

**Table 10. OLS $\hat{\beta}$ from regression of $S_t'$ against $S_t$, across spreads, $T=124$**

<table>
<thead>
<tr>
<th>Spread $S_t'$</th>
<th>Mean $\hat{\beta}$ (s.e.)</th>
<th>t-ratio against 1</th>
<th>Quartiles: median</th>
<th>0.975 (1-right tail)</th>
<th>0.05 (left tail)</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S^2$</td>
<td>1.479 (0.642)</td>
<td>0.746</td>
<td>1.378</td>
<td>2.977</td>
<td>0.637</td>
<td>0.550</td>
<td>0.463</td>
</tr>
<tr>
<td>$S^5$</td>
<td>1.475 (0.614)</td>
<td>0.774</td>
<td>1.370</td>
<td>2.918</td>
<td>0.664</td>
<td>0.569</td>
<td>0.468</td>
</tr>
<tr>
<td>$S^6$</td>
<td>1.402 (0.528)</td>
<td>0.761</td>
<td>1.335</td>
<td>2.584</td>
<td>0.644</td>
<td>0.558</td>
<td>0.457</td>
</tr>
<tr>
<td>$S^9$</td>
<td>1.340 (0.448)</td>
<td>0.759</td>
<td>1.309</td>
<td>2.288</td>
<td>0.662</td>
<td>0.570</td>
<td>0.478</td>
</tr>
<tr>
<td>$S^{12}$</td>
<td>1.263 (0.354)</td>
<td>0.745</td>
<td>1.252</td>
<td>1.977</td>
<td>0.699</td>
<td>0.619</td>
<td>0.540</td>
</tr>
</tbody>
</table>
Table 10 presents results across maturity horizons for different $S^n$. Similarly to our results for the PFS regression test in table 2, the small sample bias in $\hat{\beta}$ tends to fall when the long rate maturity horizon, $n$, is larger. For $S^2$, the mean $\hat{\beta}$ is 1.48 and it falls for larger $n$ (e.g., for $S^{12}$, $\hat{\beta} = 1.26$). Thus, the EH may be rejected, at the 97.5% level, for $\hat{\beta} = 2.9$ for some spreads (e.g., for $S^9$) while it may not be rejected, for the same value, for other spreads (e.g., $S^3$).

### 4.2.4 The Kalman filter-time varying parameter model

One of the explanations for possible empirical failures of the EH is that especially when there are regime changes, parameters may be unstable rather than constant as it is assumed by OLS. In ‘learning models’, instead of rationally knowing a constant parameter model, agents are updating the model’s parameters as they gain more information through time. In this last round of Monte Carlo experiments we consider the properties of a ‘time varying parameter’ model (TVP). We find that TVP models are an interesting case of extreme small sample bias. Although these models help eliminate the bias from regime shifts, they are considerably more exposed to persistence bias than the OLS estimator. We generate data under the null (EH), exactly as in the previous experiments, for an AR(1) short rate and a long rate constructed according to the EH. We then estimate, in each Monte Carlo trial, a time varying parameter model\footnote{We use Gauss code written by Kim and Nelson (1999) for Kalman filter estimation.} of the same PFS equation (4) that was estimated by OLS in section (4.2.1). For a more thorough description of the model, see Kim and Nelson (1999). The two main equations of the ‘state-space’ form are:

\[
S^n_t = \alpha_t + \beta_t S^n_{t-1} + \varepsilon_t, \quad \text{(Measurement equation)} \tag{3.23}
\]

\[
\begin{bmatrix}
\alpha_t \\
\beta_t
\end{bmatrix}
= \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
\alpha_{t-1} \\
\beta_{t-1}
\end{bmatrix}
+ \begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}, \quad \text{(Transition equation)} \tag{3.24}
\]

where $S^n_t$ is again the ‘perfect foresight spread’ (PFS). Given that the EH is true, we generally expect that $\hat{\beta}_t = 1$ and $\hat{\alpha}_t = 0$, for most $t$. We estimate the two ‘hyper-parameters’ $\sigma^2_{\alpha}$ and $\sigma^2_{\beta}$ and the variance of the error term in the measurement equation, $\sigma^2_{\varepsilon}$. Table 11 reports results from 5,000 Monte Carlo replications for a sample of 124 observations and for $\rho$ rising from 0 to 0.99.
Table 11. Time varying parameter PFS regression, $S_i^{(12)}$: Sensitivity to $\rho$, $T=124$

The time varying constant and slope estimate are averaged over all replications and all observations ($= T$).

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Mean $\sigma_{\epsilon}$ (PFS equation)</th>
<th>Mean $\sigma_{\eta_1}$ (('a_i' equation')</th>
<th>Mean $\sigma_{\eta_2}$ (('b_i' equation')</th>
<th>Mean $\alpha_t$ (s.d.) (t- ratio)</th>
<th>Mean $\beta_t$ (s.d.) (t- ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$1.0 \times 10^{-8}$ (-0.004)</td>
<td>0.054 (14.581)</td>
<td>$6.0 \times 10^{-10}$ (0.009)</td>
<td>0.001 (0.641)</td>
<td>1.045 (0.020)</td>
</tr>
<tr>
<td>0.50</td>
<td>$2.0 \times 10^{-7}$ (0.001)</td>
<td>0.079 (25.234)</td>
<td>$8.0 \times 10^{-6}$ (0.824)</td>
<td>-0.013 (1.187)</td>
<td>1.136 (0.032)</td>
</tr>
<tr>
<td>0.75</td>
<td>$6.0 \times 10^{-10}$ (-0.001)</td>
<td>0.156 (19.160)</td>
<td>$4.0 \times 10^{-4}$ (3.698)</td>
<td>-0.006 (2.103)</td>
<td>1.383 (0.065)</td>
</tr>
<tr>
<td>0.85</td>
<td>$1.0 \times 10^{-7}$ (9.0 \times 10^{-5})</td>
<td>0.223 (14.459)</td>
<td>0.002 (4.204)</td>
<td>0.021 (3.045)</td>
<td>1.766 (0.131)</td>
</tr>
<tr>
<td>0.93</td>
<td>$1.0 \times 10^{-9}$ (4.0 \times 10^{-3})</td>
<td>0.305 (13.000)</td>
<td>0.001 (1.541)</td>
<td>-0.014 (4.548)</td>
<td>2.710 (0.232)</td>
</tr>
<tr>
<td>0.96</td>
<td>$2.0 \times 10^{-9}$ (0.001)</td>
<td>0.378 (12.551)</td>
<td>0.006 (2.194)</td>
<td>0.001 (6.500)</td>
<td>4.641 (0.549)</td>
</tr>
<tr>
<td>0.99</td>
<td>$9.0 \times 10^{-8}$ (0.008)</td>
<td>0.500 (11.220)</td>
<td>0.159 (4.071)</td>
<td>-0.162 (10.392)</td>
<td>14.459 (3.094)</td>
</tr>
</tbody>
</table>

Considering overall results from Kalman filter estimation, for all $\rho$, the average $\sigma^2_{\epsilon}$ (over 5000 replications) is near 0. The average hyper-parameter that corresponds to $\alpha_t$, $\sigma^2_{\eta_1}$, is relatively large, implying a very high ‘signal to noise ratio’ ($\sigma^2_{\eta_1}/\sigma^2_{\epsilon}$), compared to $\sigma^2_{\eta_2}$, which corresponds to $\beta_t$. Nevertheless, the later is also large enough to imply a high ‘signal to noise’ ratio, for all $\rho$ (except at $\rho = 0$).

Turning now to the mean levels and paths for the two TVPs, $\hat{\alpha}_t$ and $\hat{\beta}_t$ (in table 11), implied by the Kalman filter algorithm and by the hyperparameter estimates, we find the same pattern of sensitivity to $\rho$ as in the case of the OLS test of the EH (in section 4.2.1). The time varying constant term $\hat{\alpha}_t$ is on average near 0 but the mean time varying slope coefficient $\hat{\beta}_t$ rises from 1 towards much higher values, as $\rho$ rises. The standard deviation around these mean estimates for $\hat{\alpha}_t$ and $\hat{\beta}_t$ rises as $\rho$ also rises. More importantly, the results suggest that the upward bias in $\hat{\beta}_t$, in the time varying parameter model, is impressively strong compared to that found for OLS estimation in section 4.2.1 (e.g. for $\rho = 0.96$, $\hat{\beta}_t = 4.64$ against 1.63 in OLS results).
This is not surprising given that, in Kalman filtering the sample used for estimation is essentially smaller than 124 observations. In the ‘basic filter’\textsuperscript{10} algorithm, information until time \( t \) is used, which is before the end of the sample, \( T(=124) \), and it is gradually updated at every time increment. As a result, the sample size is smaller than 124 at each \( t \). Moreover, the Kalman filter algorithm can be interpreted as a ‘weighted least squares’ estimator, attaching smaller weights to observations that are farther back in time. Therefore, even when the last observation \( T \) is reached, the time varying coefficient is estimated using a smaller (than 124) ‘window’ of observations. Kalman filter estimates are more comparable to those from an OLS estimator that uses a sample smaller than 124 observations.

In a time varying parameter framework, the average values and the table are only informative when complemented by the corresponding plots of average coefficient estimates for every \( t \). The properties of the estimated time varying parameter model are best summarised in the four-window plot in figure 6. The upper two windows display the plots of the ‘forecast error’ and its associated ‘conditional variance’. The forecast error\textsuperscript{11} is defined as:

\[ v_{t,t-1} = y_t - \hat{y}_{t,t-1} = S_t^2 - \hat{\alpha}_{t,t-1} - \hat{\beta}_{t,t-1} S_t \]

and the conditional variance of the forecast error is

\[ f_{t,t-1} = \mathbb{E}(v_{t,t-1}^2) \]
The lower two windows in figure 6, display the plots of coefficients as deviations from their theoretical values (\( \hat{\alpha}_t = 0 \) and \( \hat{\beta}_t = 1 \)) (i.e. estimates of the bias)\(^{12}\). Figure 6 shows that for \( \rho = 0.94 \), \( \hat{\beta}_t - 1 = 2.4 \), within a small band of 2 standard errors, while \( \hat{\alpha}_t = 0 \). The null hypothesis would have been rejected based on standard inference.\(^{13}\)

\(^{12}\) The ‘standard error bands’ display the standard deviation of the distribution obtained from the Monte Carlo trials.

\(^{13}\) We also find that \( \rho = 0.99 \) is a limiting case. The resulting path of the time varying coefficients is unusual and the properties of the time varying model are no longer robust. The time varying slope coefficient, \( \hat{\beta}_t \), rapidly rises over time to 15. The TVP model is not applicable when \( r_t \) is a near- random walk and the EH is true.
We also investigate larger samples of 524 and of 2,000 observations for this model and find that the average Monte Carlo $\hat{\beta}_t - 1$ for samples of 524 and 2000 observations (figure 7) and for a fixed $\rho = 0.928$ is always equal to 2 as the sample size increases. Unlike OLS, the TVP model is not a consistent estimator, given that, at each $t$, the Kalman filter effectively uses a window of observations for the estimation of $\hat{\beta}_t$, so that raising $T$ will not eliminate the bias.

5. Conclusions

We have used Monte Carlo Simulation experiments to explore the small sample distribution for a number of tests of the EH. Our results confirm that the small sample distribution for all these statistics is more and more distorted in response to rising persistence in interest rates and the spread. For all these tests, we also find that maturity horizon matters for small sample bias and in most cases, the latter is smaller for tests with longer horizon spreads ($S^{12}$). In accordance with results in Bekaert et al. (1997) we find strong positive bias in the OLS slope coefficient of the PFS test and a small sample distribution that has greater dispersion and is skewed to the right. The slope bias is almost eliminated when we run the same regressions without a constant term.

One of the most striking results is that the variance bound ratio, $s.d.(S_t)/s.d.(S_t^*)$, is biased downwards, for high $\rho$, hence it tends to ‘favour’ the EH too often. This is in contrast with the
OLS test that tends to reject the EH ‘too often’ and with previous Monte Carlo evidence in the variance bounds literature, that deals though with different formulations of the bound (Flavin, 1983).

Further Monte Carlo evidence shows that all the VAR methodology statistics are biased with the exception of $\text{Corr}(S_t, \dot{S}_t)$. Small sample inference would be preferable in the case of the standard deviation ratio that exhibits wide dispersion in its small sample distribution. However, the case of the Wald test statistic stands out, given that asymptotic inference is practically rendered ineffective when persistence is high. The Wald test strongly tends to over-reject the null hypothesis in small samples, as $\rho$ rises. We also look at the OLS slope coefficient in the regression of $\dot{S}_t$ against $S_t$, which, to our knowledge, is not considered elsewhere, and find a relatively high positive bias.

The Kalman filter time varying parameter model is not a typical test of the EH, however, it belongs to a broader class of learning models that may be more appropriate for testing the EH in samples which display important regime shifts. At the same time, it provides an appealing example of more extreme small sample bias (against that found for OLS) as demonstrated by our Monte Carlo evidence. Moreover, these results are relevant to a wider class of econometric techniques that may be applied to testing the EH or other theoretical models. These would include weighted least squares estimation, recursive estimation and ‘rolling window’ estimates, all of which effectively use a smaller window of observations than OLS (for the same sample).

The results presented in this paper aim to provide a guide to practitioners wishing to test the EH in markets that display high interest rate autocorrelation. Their coefficient estimates are more likely to be comparable to the quartile levels documented in our tables than to the standard critical values of asymptotic inference. That being said, Monte Carlo simulation would enhance estimation and inference if new critical values were used, derived from data generated with parameter calibration to the specific market. Finally, our results also serve as a theoretical investigation of the properties of these tests. It is hoped that Monte Carlo evidence of differences in the extent of small sample bias among alternative tests can help explain the discrepancies found in the empirical literature when different tests are employed. Towards the same direction, an investigation of the differences in the power of these tests, with data generation under a specified alternative to the null hypothesis, would provide a possible extension to our results.
References