

# Where (and by How Much) Does a Theory Break Down? With an Application to the Expectation Hypothesis

KARIM M. ABADIR and CHRISTINA ATANASOVA\*

## ABSTRACT

We provide new evidence in favor of the expectation hypothesis (EH) as a long-run theory of the term structure of interest rates. Using nonparametric techniques first, we show that the results of conventional tests that reject EH are strongly affected by the presence of extreme observations – only a handful in the case of longer maturities. We then provide a new general methodology that determines the number of outliers causing *any* theory to fail, and our approach quantifies the extent of this failure.

*JEL classification:* E4, G12, C12

*Keywords:* Expectation hypothesis, term structure of interest rates, peso problem, outliers, nonparametric estimation, recursive estimation.

\*Karim M. Abadir, Imperial College Business School, Imperial College London, UK. Christina Atanasova, Beedie School of Business, Simon Fraser University, Canada. We thank Andrea Buraschi and Sadayuki Ono. Support from ESRC grant RES062230790 is gratefully acknowledged.

# 1 Introduction

The expectation hypothesis (EH) has been the single most widely used theory of the term structure of interest rates. In its simplest form (also called the “pure” form of the EH), Fisher (1896) states that “the investor who holds a bond for a long time realizes an interest which is an “average” of the oscillating rates of those who speculate during the interim”; an attractively clear and simple statement, also quoted by Longstaff (2000b). We will focus on a weaker version of the EH that allows for a risk premium that varies with the bond’s maturity, but not with time. As the quotation indicates, the EH is a theory that has been developed with the long-term yield in mind, but we will also consider other maturities in investigating how it fares.

Different versions of this theory have been expressed in terms of spot interest rates or forward rates. But time and time again, the EH has been rejected empirically. The fact that the EH appears to be at odds with the data has spurred an enormous academic debate and a proliferation of studies that try to explain the calculated deviation from the hypothesis.

Motivated by this debate, we introduce a new general methodology to quantify where any theory breaks down, and by how much. Using this approach, we re-examine the empirical evidence on the EH and show that the failure of the conventional tests of the EH is largely due to a small number of outliers. These extreme observations correspond to a couple of rare Peso-type episodes that were typically driven by exceptional political events rather than the result of changes to economic fundamentals. These few extremes arose from an exceptionally-sharp jump in the slope of the yield curve

at a time when interest-rate levels and volatility were high, and additional considerations were at work in forming expectations. Once these outliers are accounted for with our new methodology (the outliers are *not* discarded), there is a dramatic change in the estimates of the term-structure regressions, bringing them closer to the predictions of the EH for most of the data. Our results for the EH show that, in general, the number of such extremes is very small for all but short run maturities.

The remainder of this paper is organized as follows. Section 2 introduces the EH, discusses the testable restrictions that it implies, and reviews related previous empirical results. Section 3 presents the data and discusses the usefulness of the exploratory device of nonparametric estimation in the presence of outliers and nonnormality. In Section 4, we introduce a method that determines the number of extreme observations causing any theory to fail, and our approach measures the extent of this failure, then we apply this to the EH. Instead of starting with an abstract general procedure then applying it to the EH, we opted for the reverse for ease of exposition. The final section concludes the paper.

## **2 The Expectation Hypothesis: implied relations and previous findings**

Let  $y_{m,t}$  denote the yield at time  $t$  of a zero-coupon bond maturing at time  $t + m$ , so that  $m$  is the time to maturity. Then, a linearized version of the

EH is

$$y_{m,t} = \bar{\pi}_m + \frac{\sum_{j=1}^m E_t(y_{1,t+j-1})}{m} \quad (1)$$

where, as shown in Campbell (1986),  $\bar{\pi}_m$  is a constant risk premium (or term premium) that depends on  $m$  but not  $t$ , and  $E_t$  is the expectation taken at time  $t$ . We see from this statement of the EH that the  $m$ -period rate is the average of a series of one-period expectations of future rates (recall the quotation from Fisher), plus a constant risk premium  $\bar{\pi}_m$  (which is zero in the quoted pure form of the EH). In this setup, changes in yields can only be explained by changes in the expectation of the future interest rates. On the other hand, if the risk premia were time-varying, yields would also depend on changes of attitudes towards risk or changes in interest rate volatility. Inter alia, Fisher and Gilles (1998) and Longstaff (2000a) show examples of no-arbitrage economies that support weak forms of the EH.

Interest rate persistence can generate spurious results in term structure regressions. To reduce this possibility, most empirical tests of the EH focus on differences in interest rates or spreads, rather than levels. Two relations can be derived from (1); see Shiller (1979), Shiller, Campbell, and Schoenholtz (1983), Campbell and Shiller (1991), Campbell (1995). They are:

**Type A restrictions (change in long yields):**

$$E_t(y_{m-1,t+1}) - y_{m,t} = \alpha_m + \beta_m \frac{y_{m,t} - y_{1,t}}{m-1}, \quad (2)$$

**Type B restrictions (average change in short yields):**

$$\sum_{j=1}^{m-1} \left(1 - \frac{j}{m}\right) E_t(y_{1,t+j} - y_{1,t+j-1}) = \delta_m + \gamma_m (y_{m,t} - y_{1,t}), \quad (3)$$

where  $\beta_m = \gamma_m = 1$  for all  $m$  if the EH is satisfied. Equation (2) relates future changes in long-term interest rates (the left-hand side, LHS) to the slope of the yield curve (the fraction on the right-hand side, RHS) after correction for a constant term premium. Equation (3) represents the effect of the yield spread ( $y_{m,t} - y_{1,t}$ ) on a Bartlett-type average of future changes in short-term interest rates (the LHS) which we can rewrite as

$$\sum_{j=1}^{m-1} \left(1 - \frac{j}{m}\right) E_t (y_{1,t+j} - y_{1,t+j-1}) = \frac{1}{m} \sum_{j=1}^m E_t (y_{1,t+j-1}) - y_{1,t}.$$

Testing restrictions (2)–(3) on US data typically results in strong rejections of the EH, especially for Type A and for maturities around one year or longer; e.g., see Campbell (1995, p.139). These rejections are so bad that  $\beta_m$  is found to be significantly negative, instead of being positive and equal to 1. The usual implementation of (2)–(3) replaces the expectations by their actual values, adding an unsystematic error term to the RHS, thus assuming that Muth-type rational expectations hold in addition to the EH. Froot (1989) and Balduzzi, Bertola, and Foresi (1997) find that expectational errors are partly to blame for the failure of this tested version of the EH. In particular, focusing on Type A restrictions, Froot (1989) shows that, expectational errors apart, the EH holds at long maturities but not at short ones; see also Table 5 in Rossi (2007). Furthermore, using additional data on overnight rates, Longstaff (2000b) shows that the EH holds at very short maturities. So what explains the failure of Type A restrictions around one-year maturities?

The most widely used methods for testing restrictions (2)–(3) is least squares (LS) regression. Other methods of parametric estimation have also been used, but the results are little changed; e.g., see the instrumental-

variable estimation in Campbell and Schiller (1991). We will therefore focus on the LS approach and its local versions.

Suppose that we wish to predict a variable  $z$  by using an optimal function  $h(x)$  of another variable  $x$ , with optimality meaning that we choose  $h$  to minimize the expected squared error  $E(z - h(x))^2$  as per the LS criterion. It has been proven that this is mathematically equivalent to the calculation of the conditional expectation  $E(z | x)$  (which is a function of  $x$ ) in the generic “regression” setup

$$z = E(z | x) + \varepsilon. \quad (4)$$

In fact, one way of *defining* a conditional expectation is to write it as the  $h(x)$  that minimizes  $E(z - h(x))^2$ , with  $E(z | x)$  obtained as the solution for the optimal function  $h(x)$ ; e.g., see Grimmett and Stirzaker (2001, p.346) or Abadir, Heijmans, and Magnus (2018, p.217). Now, conditional expectations  $h(x)$  are linear in  $x$  for the class of elliptical distributions for  $(x, z)$  (a justification for linear regressions), which includes the normal and Student t, but they are nonlinear for the others. Visually, elliptical distributions give rise to scatterplots that look like shaded ellipses.

A scatter plot of  $(x, z)$  in the case of our restrictions (2)–(3) reveals that the bivariate density deviates from elliptical shapes at the edges of the dataplot, as we shall see in the figures below, so instead of restricting the functional form of the conditional expectation to linearity we allow the data to speak for themselves by means of nonparametric (NP) estimation in order to see where the problems are. More specifically, the Nadaraya-Watson regression is the estimator of  $E(z | x)$  for (4), arising from the usual NP kernel estimator of the joint density of  $(x, z)$ ; e.g., see Li and Racine

(2006) for the definition of these estimators.<sup>1</sup>

The NP estimation can reveal the functional form of the conditional expectation and how far (and where) it deviates from linearity and from  $\beta_m = \gamma_m = 1$ . Being a local-LS device, the Nadaraya-Watson regression also shows the extent to which earlier results in the literature were distorted by a handful of extreme observations driven by unusual events: in 1981/1982 (the Volcker experiment) and 1987 (the stock market crash). These events are the main cause of the violation of the elliptical form of the bivariate densities, which we will also be able to see in the scatterplots. Thus, in addition to being robust to outliers, our results are robust to nonnormality as stated earlier in the motivation for NP.

Froot (1989, p.287) gives a brief Peso-problem interpretation for the violation of the EH. Another empirical study supporting the Peso problem explanation is Evans (1996). Bekaert, Hodrick, and Marshall (2001) develop a test statistic and find empirical evidence against the EH when the risk premium is constant. However, they suggest that an important reason for the rejection of the EH is a Peso problem due to the inflation process. We note that, historically, the US economy has rarely been in a high-inflation regime, and we will see later from our analysis that the main violations of the EH are contained within such rare periods. Fama's (2006) result supports the idea that there was a problem (for the EH) caused by the peaking of the short

---

<sup>1</sup>It is possible to use asymmetric kernels in this context, since we suspect the direction of the asymmetry; see Abadir and Lawford (2004) and Bouezmarni and Scaillet (2005). However, we take an alternative route to improving efficiency after the initial exploratory NP regression, and our approach is detailed in Section 4.

rate in the early 80's, but our analysis of the impact of this period on the regressions will give different interpretations for the violation of the EH.

### **3 Nonparametric evidence: where do the violations of the EH come from?**

Our data consist of monthly yields on zero-coupon Treasury bonds, extracted from coupon bond prices using the smoothed Fama-Bliss estimation method (see Fama and Bliss, 1987, for details). The interest rate data have to be estimated because the zero-coupon Treasury bonds are available for 3-month, 6-month and 12-month maturities, while Treasury bonds exceeding a 12-month maturity usually pay coupons. The monthly data cover the period from January 1959 to December 2000, and the total number of observations is 503 months. We present empirical results for zero-coupon bonds with constant maturities of 3, 6, 12, 24, 48, 60, and 120 months.

As discussed in the previous section, the most striking violation of the EH is for the Type A restriction, the test results being worse as maturity increases. The violations of the Type B restriction are not as bad because its LHS in (3) is an average and, as we show in the next section, is less affected by occasional outliers. The estimates of Type B regressions have positive slopes  $\hat{\gamma}_m$ , unlike Type A regressions where the estimated slopes  $\hat{\beta}_m$  are negative and large.

We now present graphs of NP regressions, where we also include the linear least squares fit for comparison. For the sake of brevity, we only report two

representative maturities around the problematic one-year maturity for Type A of (2). In each of the two cases, we present two graphs for the *same* NP regression: one shows all the data points, while the other zooms into the same picture to show in more detail the fitted NP regression curve for the bulk of the data.

In Figure 1, the NP regression for  $m = 12$  months looks like a flat line, except for a handful of outliers at each extremity. But the scales on the two axes are very different: the vertical axis is squashed as a result of visualizing all the datapoints, and it is difficult to see where and how much the slope differs from 1. To do this, we zoom in on the body of the graph and represent this in Figure 2. For the x-axis ranging from about 0.05 to 0.14, the fit is amazingly linear for a nonparametric regression (this linearity has not been imposed at the outset) with a slope of about 0.7. Even when taking all the points on this graph, the “average” slope is little changed. This is in sharp contrast to the large negative number obtained by LS regression, namely  $-1.2$  with a standard error (SE) of 0.5 (or adjusted to 0.6 when we account for heteroskedasticity and 0.7 accounting for autocorrelation as well) leading to the rejection of the EH. We can now see that the puzzling LS result was due to the handful of extreme values at the right in Figure 1 that are driving the “average” slope down, and that the slope for the bulk of the data is statistically indistinguishable from the unit coefficient that is implied by the EH. A similar story repeats in Figures 3 and 4 for the 2-year maturity,  $m = 24$ , albeit with a larger SE of 0.6, or 1.0 when adjusted for heteroskedasticity (Newey-West adjusting for autocorrelation has no further effect). In regressions of Type A as (2), the spread is divided by  $m$ , hence

our reporting of the horizontal axis in Figure 4 ( $m = 24$ ) being half that of Figure 2 ( $m = 12$ ).

The graphs for Type B are not as dramatically different for the two approaches, compared to the visual difference we see for Type A. Therefore, we do not report Type-B graphs for space considerations. For typical values of the spreads in the range of  $-0.5$  to  $2.0$ , the average slope of the NP regression is approximately  $0.4$ , versus  $0.3$  to  $0.4$  for LS with SEs of  $0.1$  whether or not we correct for heteroskedasticity. Apart from the slopes being positive, one noticeable difference here is that correcting Type B regressions for autocorrelation using Newey-West estimators, the SEs double to  $0.2$ . This is because the LHS of Type B regressions (3) is an average that induces the overlap of observations, hence the presence of some serial correlation for such a specification of the EH equation.

Finally, we note that the data points that generated these large outliers for the spread are mostly in 1981–1982, and also the last two months of 1987, the months that followed the famous October crash. Consistent with the Peso interpretation of the previous section, 1981–1982 was a period of a severe recession in the US, which was caused by a contractionary monetary policy established by the Fed to control high inflation. In the next section, we quantify more rigorously how many such observations cause the rejection of the EH.

## 4 The last straw: how many observations cause the rejection of the hypothesis?

In this section, we introduce our parametric regression methodology that determines at which point the parameter ceases to be consistent with economic theory. This is done by means of two types of regression, in increasing order of complexity in the first and second subsections: Recursive LS (RLS) and LS with sequential thresholds, respectively. The third subsection discusses the methodology in a more abstract setup than the EH.

### 4.1 RLS on the sorted data

In the notation of (4), we start by sorting the sample's data pairs  $(x, z)$  in increasing values of  $x$ : from the most negative to most positive spread. Using RLS on the spread-sorted data, we estimate  $\beta_m$  and  $\gamma_m$  as we gradually enlarge the sample size from the first 100 pairs  $(x, z)$  with the smallest  $x$ , until we eventually include all the data. We do this to determine formally which data points and corresponding events cause the failure of the EH. The point where this happens is when the band of  $\pm 2\text{SE}$  around the parameter estimates does not include the value  $\beta_m = 1$  or  $\gamma_m = 1$ . We use heteroskedasticity-consistent SEs because the sorting itself induces heteroskedasticity, in addition to any that may have been present in the unsorted data. However, the sorting destroys autocorrelations because they depend on the ordering on the data, and these can be important in Type B regressions as noted in the previous section. We will deal with this aspect in the next

subsection.

**Table I**

Percentage of high extremes in the spread that make

the EH fail in Type A or B

| $m$ : Bond maturity (months) | 2  | 3  | 6  | 12 | 24  | 48  | 60  | 120 |
|------------------------------|----|----|----|----|-----|-----|-----|-----|
| % for EH to hold in Type A   | 22 | 21 | 19 | 2  | 0.4 | 0.4 | 0.4 | 2   |
| % for EH to hold in Type B   | 17 | 22 | 42 | 36 | 29  | 7   | 5   | 3   |

In Table I, we report the percent of the sample observations that triggers the rejection of the EH. We see that the EH performs better at longer maturities, which is unsurprising given that the EH is a theory that is designed with the long run in mind; e.g. see the statement in our introduction. What is surprising, however, is that in many cases, only few observations are causing the rejection: only 2 observations (0.4% of the sample) for  $12 < m < 120$  in the case of Type A. Perhaps this is most dramatically illustrated by the RLS graph for  $m = 24$  for the Type A regression, in Figure 5, where the vertical axis measures the estimate  $\hat{\beta}_m$  (middle curve) and its 95% confidence bands,  $\hat{\beta}_m \pm 2SE$ , the horizontal axis representing how many data points having the smallest spread are included. The horizontal line denoting  $\beta_m = 1$  is crossed after 501 sorted observations, indicating that the EH is violated only because of the largest 2 spreads in the dataset of 503 observations, both data points being in 1982.

The explanation for the higher percentage for Types B than A in Table I has already been alluded to, here and in the previous section. The LHS

of Type B is an average, and this average therefore carries these outliers in more than just a couple of time periods. Furthermore, because of the overlap of the observations in contiguous averages, an autocorrelation is induced by the averaging, and it requires adjusting the SEs. This cannot be done in the sorted data, but we now tackle this in the original data by using a more elaborate regression specification.

## 4.2 LS with sequential thresholds, on the original data

Instead of regressing  $z_t$  on  $x_t$  as we did before (using the generic terminology  $x, z$  for the variables of regressions of Types A & B), we create a dummy variable  $D_{i,t} = 1$  when  $x_t \leq i$ , and  $D_{i,t} = 0$  otherwise. We repeat this regression for the following sequence of thresholds: the first regression starts with the largest possible threshold (equal to the largest spread), then we reduce the threshold by 0.001 (alternatively, by one data point) for each subsequent regression, stopping when the EH is not rejected anymore. The new equation to be estimated (for each  $m$ ) can be written as

$$z_t = a_i + b_i (D_{i,t} x_t) + c_i ((1 - D_{i,t}) x_t) + \varepsilon_{i,t}, \quad (5)$$

where we test  $b_i = 1$  using heteroskedasticity and autocorrelation consistent (HAC) SEs.<sup>2</sup> The component  $c_i ((1 - D_{i,t}) x_t)$  is where we evaluate the effect of extreme spreads on the regression and on deviations from the EH. For Type B regressions, longer maturities imply more observations in the LHS

---

<sup>2</sup>Our methodology is not restricted to the use of HAC: other types of corrections to SEs can be used here. For example, Robinson's (2005) MAC can be used; see Abadir, Distaso, and Giraitis (2009) for comparisons.

averaging, hence more autocorrelation and a wider HAC window. Note that the alternative (instead of (5)) approach of removing the observations where  $D_{i,t} = 0$ , and regressing  $z_t$  on  $x_t$  in the remaining subsample, would destroy the structure of the autocorrelation again, if many observations are removed, in which case HAC would not play its full role because it would have no effect beyond heteroskedasticity adjustment. Note also that it would be inefficient to remove from (5) the component involving  $c_i$ , although the estimate of  $b_i$  would be unaffected because of the orthogonality of the regressors. It would also makes no economic sense because such an incomplete regression would claim that  $z_t$  is unaffected by  $x_t$  when we have an extreme observation leading to  $D_{i,t} = 0$ .

**Table II**

Percentage of high extremes in the spread that make  
the EH fail in Type A or B

| $m$ : Bond maturity (months) | 2         | 3         | 6         | 12        | 24       | 48       | 60       | 120      |
|------------------------------|-----------|-----------|-----------|-----------|----------|----------|----------|----------|
| % for EH to hold in Type A   | <i>12</i> | <b>22</b> | 19        | 2         | 0.4      | 0.4      | 0.4      | <i>1</i> |
| % for EH to hold in Type B   | <i>11</i> | 22        | <i>36</i> | <b>38</b> | <i>3</i> | <i>0</i> | <i>0</i> | <i>0</i> |

Notes: HAC SEs are used for estimating (5) on unsorted data.

Italics indicates EH improvement over Table I, boldface worsening, roman otherwise.

In Table II, we report the percentage of observations that cause the EH to break down. Comparing it to Table I, we use italics to indicate an improvement, boldface for worsening, and roman otherwise. There are only two

cases where the results are worse and, even so, the deterioration is marginal (by 1% and 2%); but the cases with improvements can be substantial. The best improvements are for Type B at maturities longer than a year (as expected from the averaging argument given earlier in this section), but also surprisingly at the short end. The improvements are substantial at  $m = 2$  for Types A & B (which reinforce the findings of Longstaff, 2000b) and  $m > 12$  for Type B. The results for 3–6 months (Type A) and 3–12 months (Type B) are mixed but still high, indicating that outliers are not enough to explain the deviations from the EH in these few cases. Nevertheless, Della Corte, Sarno, and Thornton (2008) show that rejections at such maturities are economically insignificant, even if they were statistically significant.

### 4.3 Further discussion of the methodology

As we have illustrated in the previous subsection, the applicability of our approach is to an outlier-caused breakdown of a hypothesis. Although indicative, our methodology is not meant to evaluate per se other causes of failures of an economic hypothesis, such as incorrect explanatory variables or inadequate functional forms, topics which are well covered elsewhere. It is possible to alter the  $(1 - D_{i,t})x_t$  term in (5) to include remedies for other explanatory variables and/or functional forms to indicate what changes in the relation have taken place at the extremes, but the crux of the methodology would still be about finding at which extreme-point does a theory like  $z_t = a + bx_t + \varepsilon_t$  (with  $b$  a hypothesized specific constant like 1) cease to be consistent with the data. As seen in the case of a couple of maturities above,

the percentage of “extremes” was a high proportion of the sample, indicating that outliers were not the cause of the theory’s failings for these maturities.

In our application to the EH, the initial data exploration by NP regressions allowed us to determine that estimates of  $b$  were extremely sensitive to a few large positive spreads in the dataset, and that this did not happen at the opposite end of large negative spreads. It is possible to extend the results of the previous subsection to the case of extremes at both ends of  $x$ , by having an indicator function (dummy)  $1_{\mathcal{K}} = 1$  if condition  $\mathcal{K}$  holds, and zero otherwise: with two thresholds ( $i$  and  $j$ ), we generalize (5) to

$$z_t = a_{ij} + b_{ij} (1_{x_t \in [i,j]} x_t) + c_{ij}^l (1_{x_t < i} x_t) + c_{ij}^u (1_{x_t > j} x_t) + \varepsilon_{ij,t}, \quad (6)$$

with  $c_{ij}^l$  and  $c_{ij}^u$  the slope coefficients for  $x < i$  and  $x > j$ , respectively. As a grid is constructed over the two sides, there is no order in general for nesting the two sequences of grids. Instead, one evaluates the regression over the combinations of  $i, j$  within windows of containing  $I, J$  data points, respectively, then choosing the combination(s) of  $i, j$  minimizing  $I + J$ . For example, if each window contains two observations ( $I = J = 2$ ), the smallest two observations in  $I$  and the largest two in  $J$ , then the regression’s  $b_{ij}$  is evaluated for the 9 combinations that exclude: 0, 1, or 2 lower extremes, combined with 0, 1, or 2 upper extremes. The size of  $I$  and  $J$  is enlarged until there is no rejection of the hypothesis  $b = 1$ , and a minimum value of  $I + J$  is found. Note that a minimum  $I + J$  may correspond to more than one combination of  $i, j$  (for example, 1, 2 or 2, 1) and that additional criteria may be chosen by the user to select one specific  $i, j$  if this is required.

If  $x_t$  or  $z_t$  were i.i.d. (not the case for the EH data), then a crude way

to deal with extremes would be to trim the data or winsorize them. In addition to the latter removing extremes instead of modelling them, there is no guarantee that the large values that are being removed are actually extremes; for example, in our context, a large  $z_t$  may well be compatible with a large  $x_t$  and not produce an estimate of  $b$  that is different from 1. Further challenges to trimming would arise for non-i.i.d. data, as they would require pre-filtering them with a model (which is possibly incorrect) before trimming the resulting residual; for an example of a simple pre-filtering model, if a trend exists in the data then it should be removed before considering what is an extreme deviation from the trend. Uncertainty about the model that is used for pre-filtering the data adds a possibility of misspecification. Finally, trimming the data requires the choice of some arbitrary cutoff, which is not the case with our methodology where a sequence of thresholds is considered. We actually *estimate* the best (in a minimalist sense) threshold from the data.

One can regard our model as having a number of regimes; for example, (6) can be said to have three regimes in the case where all three RHS components are needed and none of them drops out. Unlike Markov-switching models, the switching is not due to the probability of some unobservable variable causing a switch into a different regime: the switching is caused by the observed outlier, and the threshold at which the switching takes place is deterministic. Our methodology is also valid in contexts that are not necessarily dynamic.

## 5 Concluding comments

We investigated the role that outliers play in the rejection of the EH, and found that they are a crucial factor leading to its rejection in all but some short maturities. We also found that, surprisingly, only a handful of observations are responsible for the large rejections seen in the literature. Once these exceptional Peso-like events are accounted for, the EH seems to hold for the rest of the time.

In obtaining our results, we introduced a new way of determining the points where a theory breaks down. This new methodology can be used in other applications beyond the current field of study of interest rates. It can be used in a static or dynamic context, not just in time series as is the case for traditional recursive estimation. Furthermore, our approach provides a testing framework based on a new estimation procedure.

## REFERENCES

- Abadir, Karim M., Walter Distaso, and Liudas Giraitis, 2009, Two estimators of the long-run variance: beyond short memory, *Journal of Econometrics* 150, 56–70.
- Abadir, Karim M., Risto D. H. Heijmans, and Jan R. Magnus, 2018, *Statistics* (Cambridge University Press, Cambridge).
- Abadir, Karim M., and Steve Lawford, 2004, Optimal asymmetric kernels, *Economics Letters* 83, 61–68.

- Balduzzi, Pierluigi, Giuseppe Bertola, and Silverio Foresi, 1997, A model of target changes and the term structure of interest rates, *Journal of Monetary Economics* 39, 223–249.
- Bekaert, Geert, Robert J. Hodrick, and David A. Marshall, 2001, Peso problem explanations for term structure anomalies, *Journal of Monetary Economics* 48, 241–270.
- Bouezmarni, Taoufik, and Olivier Scaillet, 2005, Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data, *Econometric Theory* 21, 390–412.
- Campbell, John Y., 1986, A defense of traditional hypotheses about the term structure of interest rates, *Journal of Finance* 41, 180–193.
- Campbell, John Y., 1995, Some lessons from the yield curve, *Journal of Economic Perspectives* 9, 129–152.
- Campbell, John Y., and Robert J. Shiller, 1991, Yield spreads and interest rate movements: a bird’s eye view, *Review of Economic Studies* 58, 495–514.
- Della Corte, Pasquale, Lucio Sarno, and Daniel L. Thornton, 2008, The expectation hypothesis of the term structure of very short-term rates: statistical tests and economic value, *Journal of Financial Economics* 89, 158–174.
- Evans, M.D.D., 1996, Peso problems: their theoretical and empirical implications, in G.S. Maddala and C.R. Rao, eds.: *Handbook of Statistics*, vol. 14 (Elsevier, Amsterdam).

- Fama, Eugene F., 2006, The behavior of interest rates, *Review of Financial Studies* 19, 359–379.
- Fama, Eugene F., and Robert R. Bliss, 1987, The information in long-maturity forward rates,” *American Economic Review* 77, 680–692.
- Fisher, Irving, 1896, Appreciation and interest, *Publications of the American Economic Association* 11, 21–29.
- Fisher, Mark, and Christian Gilles, 1998, Around and around: the expectations hypothesis, *Journal of Finance* 53, 365–383.
- Froot, Kenneth A., 1989, New hope for the expectations hypothesis of the term structure of interest rates, *Journal of Finance* 44, 283–305.
- Grimmett, Geoffrey R., and David R. Stirzaker, 2001, *Probability and Random Processes*, 3rd ed. (Oxford University Press, Oxford).
- Li, Qi, and Jeffrey Scott Racine, 2006, *Nonparametric Econometrics: Theory and Practice* (Princeton University Press, Princeton).
- Longstaff, Francis A., 2000a, Arbitrage and the expectations hypothesis, *Journal of Finance* 55, 989–994.
- Longstaff, Francis A., 2000b, The term structure of very short-term rates: new evidence for the expectations hypothesis, *Journal of Financial Economics* 58, 397–415.
- Robinson, P. M., 2005, Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction, *Econometric Theory* 21, 171–180.

- Rossi, Barbara, 2007, Expectations hypotheses tests at long horizons, *Econometrics Journal* 10, 554–579.
- Shiller, Robert J., 1979, The volatility of long-term interest rates and expectations models of the term structure, *Journal of Political Economy* 87, 1190–1219.
- Shiller, Robert J., John Y. Campbell, and Kermit L. Schoenholtz, 1983, Forward rates and future policy: interpreting the term structure of interest rates, *Brookings Papers on Economic Activity* 1, 173–217.

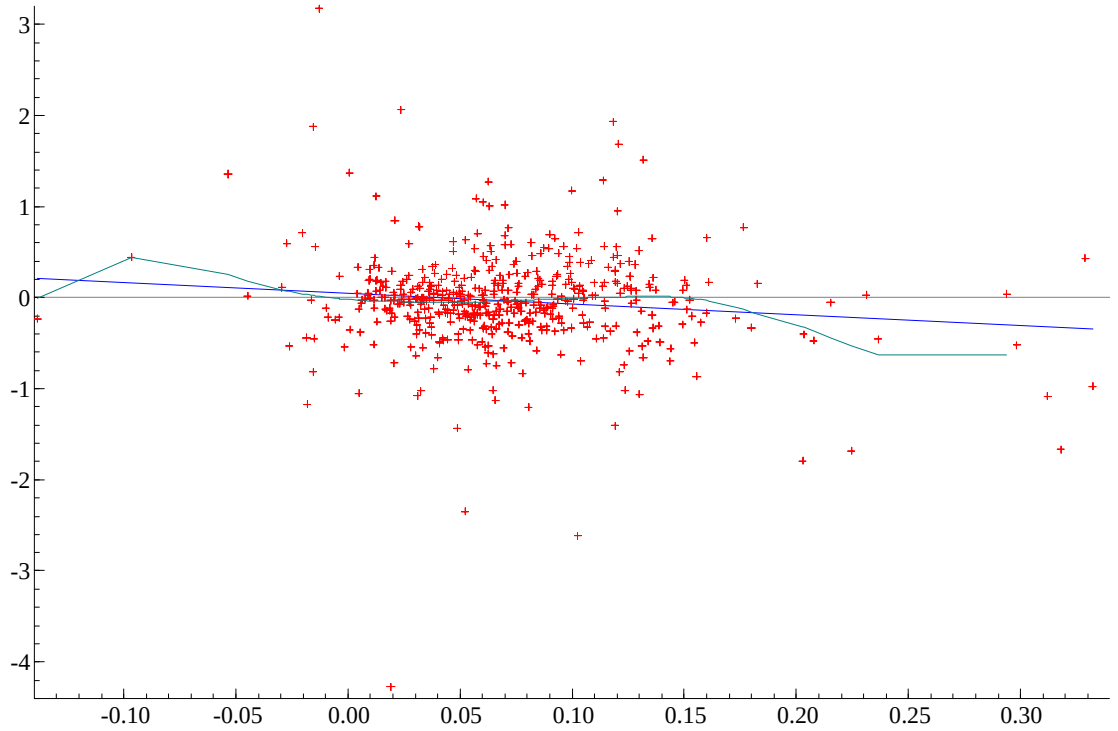


Figure 1: LS (line) and NP (curve) regressions for Type A restrictions for  $m = 12$  months, representing the change in long yields as a function of the slope of the yield curve (the spread divided by  $m$ ).

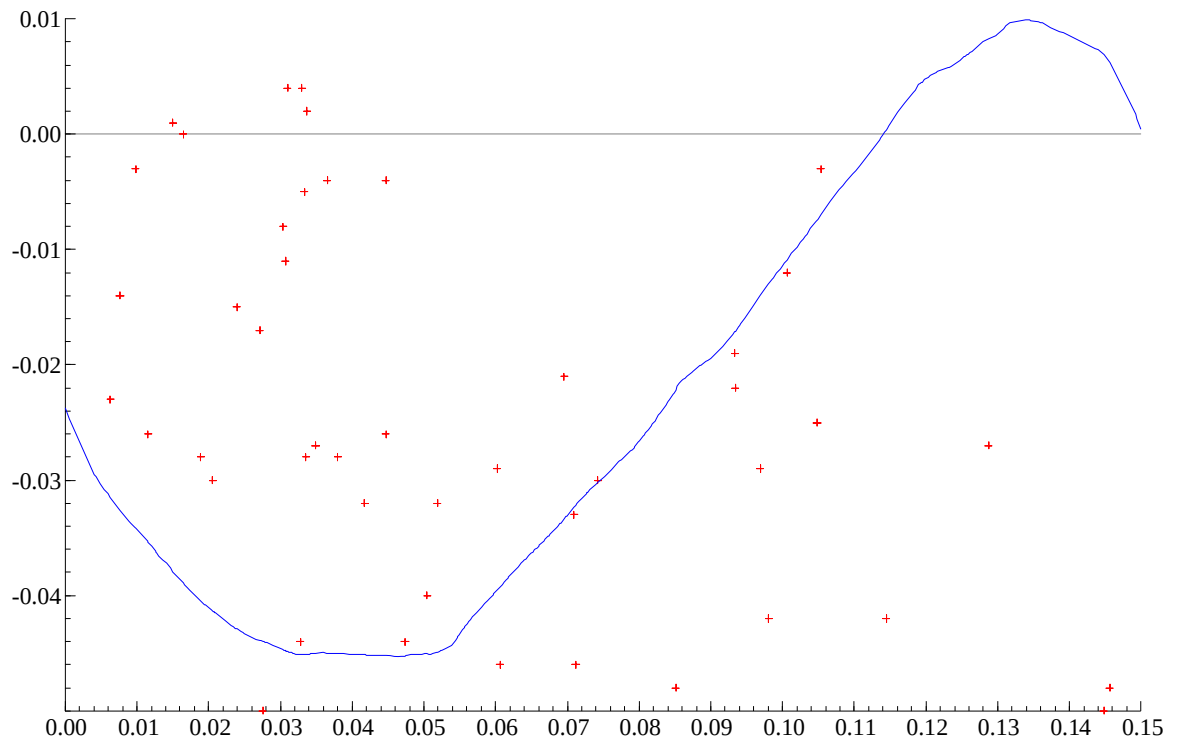


Figure 2: An enlarged picture of the NP regression of Figure 1.

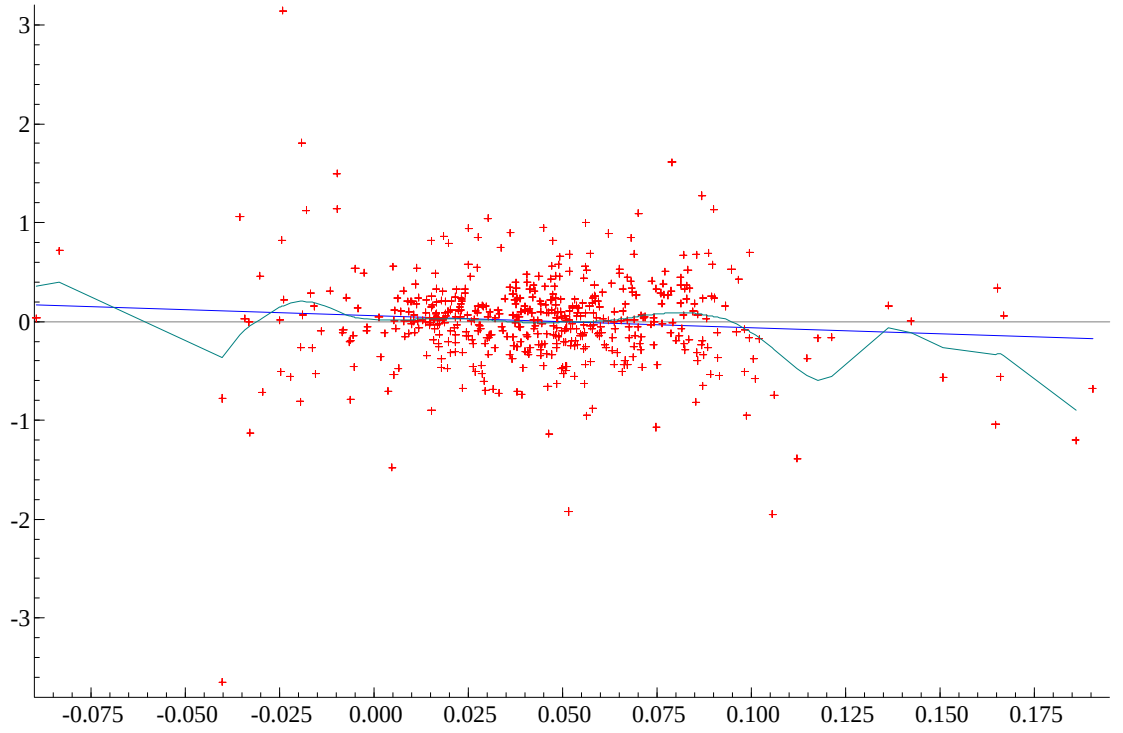


Figure 3: LS (line) and NP (curve) regressions for Type A restrictions for  $m = 24$  months, representing the change in long yields as a function of the slope of the yield curve (the spread divided by  $m$ ).

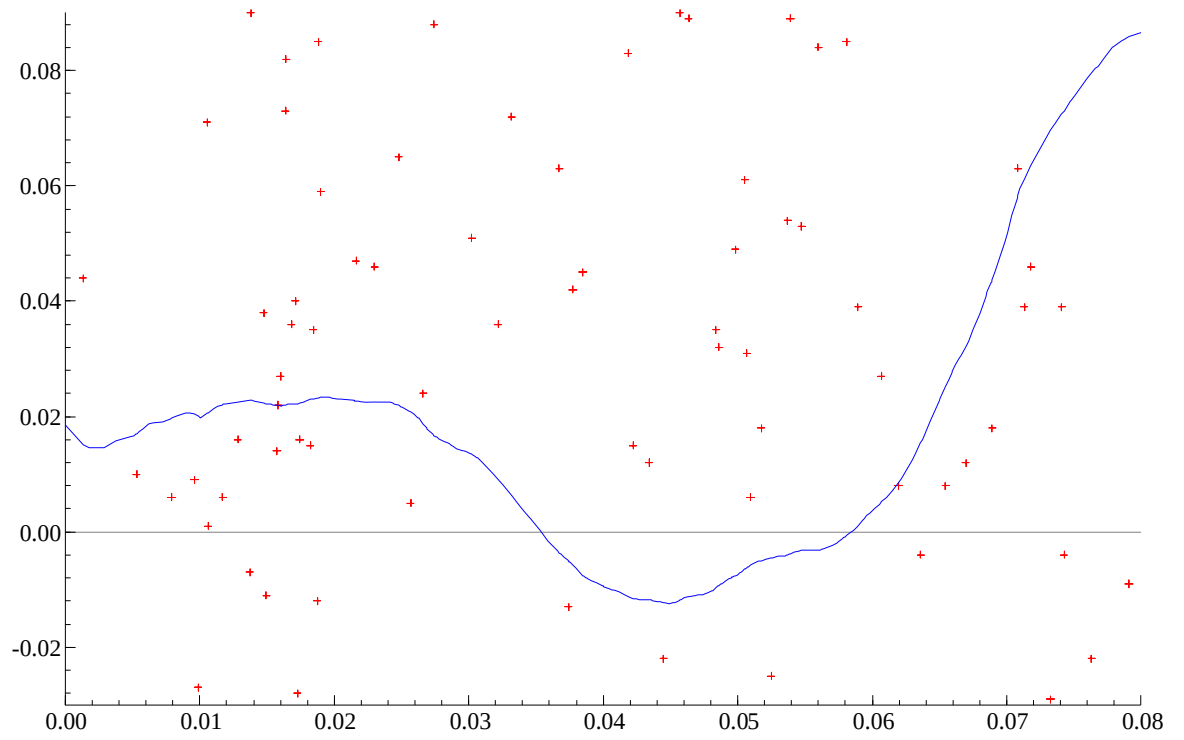


Figure 4: An enlarged picture of the NP regression of Figure 3.

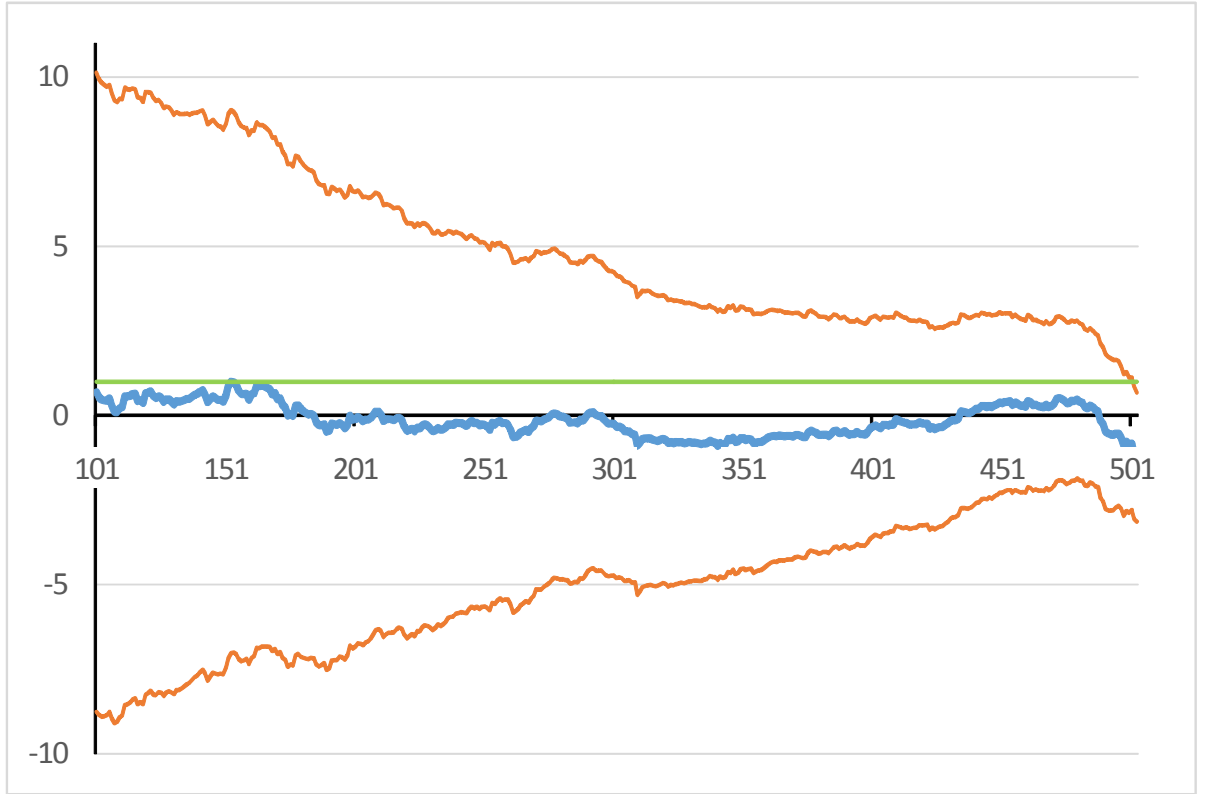


Figure 5: Recursive LS estimate (middle curve) and its  $\pm 2\text{SE}$  confidence bands, for  $m = 24$  in spread-sorted data for Type I regression. The horizontal line is  $\beta_m = 1$  and it is crossed after 501 sorted observations, indicating that EH is violated because of the largest 2 spreads in the dataset of 503 observations.