An Exact Method for Establishing Significance in Time Series Analysis with Finite Samples and Bounded Errors

(preliminary draft)

Heiko Rachinger Karl H. Schlag

December 10, 2014

Abstract

We present the first genuine nonparametric test for the autoregressive coefficient of a times series model, with the unit root test as special case, when errors have given bounds and follow a martingale difference sequence. Without such bounds nontrivial tests are known not to exist. Our test is exact, we do not add other assumptions on the process. Competitors either do not control the type I error for a given finite sample (such as the Dickey-Fuller test) or rule out outcome variables with finite support and test a more restrictive null hypothesis by also assuming that all conditional medians are equal to 0.

1 Introduction

Regression analysis is arguably the most popular tool of econometrics. Including lagged variables as explanatory variables poses a challenge to inference, as for instance unbiased estimates no longer exist. A special and very prominent class of examples that involves lagged explanatory variables are autoregressive models. An additional challenge included in the time series literature is that one wishes to relax independence and only impose errors that follow a martingale difference sequence. Unit root tests receive lots of attention, for instance they are used to assess efficiency of financial markets (for a survey see Charles and Darné, 2009). Yet richer tests are needed to understand the degree of nonstationarity in the data when the unit root cannot be rejected. After all, lack of the ability to reject a unit root may simply be due to lack of data or insufficient power of the underlying test. One needs confidence intervals to be able to uncover, when small, that data is nonstationary or close to being nonstationary.

As the importance of applications rises, and with the advancement of both theory and computing techniques, exact methods gain increased attention. To be exact means no more than to be correct for the given finite sample and the given model of the data. Tests derived using asymptotic theory, hence ignoring the limited availability of data, can lead to drastic misrepresentations of the findings. Size distortions easily rise above 30% for instance for the classic test of Dickey and Fuller (1979). We include some simulations for additional illustration.

Ideally one would like to design a test that is correct in finite samples, while only assuming that errors follow a martingale difference sequence. Most importantly it seems that one should be able to allow for heteroskedasticity. However, following Dufour (2003) we know that nontrivial exact tests do not exist without further restrictions on the data generating process. Exact tests under heteroskedasticity for the unit root were derived by Wright (2000) and Luger (2003), who both impose conditions that imply that errors have a conditional median equal to 0. This weak form of symmetry severely restricts the data generating process and seems to contrast the original motivation and interest for heteroskedastic modelling.[1] We present simulations that show that the tests of Wright (2000) and Luger (2003) can be dramatically oversized when errors are asymmetric. This raises the doubt, when evaluating one of these tests in real data sets, whether a rejection is really due to a violation of the martingale hypothesis or instead simply due to the failure of all medians being equal to 0. We do not wish to impose constraints on the median. Instead we ensure existence of a nontrivial test by assuming bounds on the errors. Unlike these other papers we do not put any constraints on possible point masses of the errors and allow for a trend.

There is an extensive literature on martingale inequalities with bounds, however we only have come across one paper (van de Geer, 2002) which highlights how one might design tests in autoregressive processes under heteroskedasticity. We improve the inequality of van de Geer (2002), and proceed differently than the proposal of van de Geer (2002).This inequality is useful when due to correlation the future is easier to predict under the alternative than under the null hypothesis. An inequality of Pinelis (2006) is built in to accommodate for situations in which the future is difficult to predict.

Bounds occur naturally in many applications, such as non negativity constraints. Bounds are a natural part of inference when one wishes to argue that past observations are typical for what should be expected in the future. Data ranges are observable and easy to describe, as opposed to imposing knife edge conditions such as the conditional median being equal to 0. One might choose to impose bounds on the dependent

[1] The assumptions made by Wright (2000) and Luger (2003) are sufficiently restrictive that they allow one to derive the distribution of the test statistic, this is not possible for our set of assumptions.

variable. We choose to impose bounds on the errors as innovations that drive the time series are the center of interest. As inference will depend on which bounds are chosen we produce joint confidence sets for bounds on the error and the value of the autoregressive coefficient.

Our approach is general as we provide a test for any value of the autoregressive coefficient, enabling the construction of confidence intervals. Our approach is general as it paves the way for including covariates and thus for the analysis of dynamic panels where there is not a clear separation between dependent and independent variables (cf. our conclusion). Software will be posted on our homepage to allow users to immediately apply our tests.

We proceed as follows. In Section 2 we present the model, Section 3 contains an overview of the literature and how we found our approach. In Section 4 we describe our test. Section 5 contains the formal details. Subsection 5.1 presents our generalization of the inequality of van de Geer (2002) and how we use it for treating stationary alternatives, in Subsection 5.2 we show how to use an inequality of Pinelis (2006) when the alternative is nonstationary, Section 5.3 then shows how to combine the two methods. In Section 6 we present simulations to show when tests of Dickey Fuller (1979), Wright (2000) and Luger (2003) fail to control type I errors. It also contains simulations that show the power of our tests and their ability to provide evidence for a unit root by establishing an exact confidence interval. In Section 7 we apply our test to real data, Section 8 contains the conclusion.

2 The Model

Consider the simple autoregressive model defined by $Y_0 = y_0$ and $Y_t = a + bY_{t-1} + \varepsilon_t$ for $t = 1, .., T$ with the following specifications. a and b are unknown, $a, Y_t \in \mathbb{R}$ and $b \in [0, 1]$. $\{\varepsilon_t\}_t$ is a martingale difference sequence given the filtration F that is defined by $\{Y_t\}$, so $E(\varepsilon_t | F_{t-1}) = 0$ for all t. Furthermore, there are some F_{t-1} measurable random variables g_{t1} and g_{t2} such that $g_{t1} < g_{t2}$ and $g_{t1} \leq \varepsilon_t \leq g_{t2}$ for all t. Let $g_t = \frac{1}{2}(g_{t2} - g_{t1})$. We call this model (a). We also consider the analogous model with the same specifications as above except there is also a time trend, so $Y_t = a + bY_{t-1} + ct + \varepsilon_t$ for $t = 1, .., T$ where c is unknown, $c \in \mathbb{R}$. This we refer to as model (ac).

Note that we do not make any explicit assumptions on the possible values of Y_t. In particular one can apply our method to the case where Y_t has finite support, for instance one may choose to investigate data in which $Y_t \in \{0, 1\}$ for all t. The latter cannot be sensibly done with any method that imposes the median error is equal to 0. Of course the boundedness of our errors implicitly defines bounds on the range of the outcome Y.

For each b_0 we will construct a test for $H_0 : b = b_0$ against $H_0 : b \neq b_0$. In particular, when $b_0 = 1$ we are presenting a test of a unit root. Having a test for $H_0 : b = b_0$ for each b_0 enables us to design a confidence interval for b with guaranteed coverage. The tests we construct will be exact in the sense that the presented type I error probabilities are correct for the given finite sample.[2] In particular we do not use asymptotic theory to derive the level of any of our tests.

3 An Overview

Existing tests fail to be exact under the typical conditions found in real data that include errors that are heteroskedastic, asymmetric and in particular not normally distributed. Their p values are not valid for these environments. In some cases this is because their properties are derived using asymptotic theory. A popular example is the test of Dickey and Fuller (1979) for a unit root which is known to be oversized when errors are heteroskedastic (see e.g. (Kim et al., 2002)). In other cases this is because the tests are only exact for more restrictive environments. Wright (2000) (with variations presented by Belaire-Franch and Contreras (2004) and Kim and Shamsuddin (2008)) constructs an exact test for the case where both the conditional mean and the conditional median are equal to 0 and errors are almost surely unequal to 0.[3] Luger (2003) presents an exact test for the special case where the joint distribution of the errors is symmetric and no two errors obtain the same value with positive probability.[4] In Section 6 we demonstrate how both of these tests can be drastically oversized when errors are asymmetric. On top of this, both papers either implicitly or explicitly restrict the possibility of having error distributions with point masses. In particular this means that they cannot be applied to data where the outcome variable has finite support such as in binary time series. We do not impose any constraints on point masses of the errors nor do we make any restrictions on the possible values of the outcome variable Y. In addition, we allow for a time trend.

In this paper we construct a test and prove that it is exact. We believe that it is instrumental to explain how we found our test, and not to simply present and prove

[2] Following Yates (1934), a test is called *exact* if its type I error probability never exceeds the given nominal level.

[3] Wright's (2000) assumptions are a bit more general than stated here. In addition to $E(\varepsilon_t|F_{t-1}) = 0$ it is assumed that $P(\varepsilon_t > 0|F_{t-1}) = 1/2$ for all t. This latter assumption is hard to motivate on an intuitive level unless one assumes that $\varepsilon_t \neq 0$ holds almost surely, in which case it is equivalent to assuming that the conditional median error is always equal to 0. This then however rules out outcome variables with finite support unless one adds unreasonable contraints on the possible values of a.

[4] Luger (2003) uses in the proof of a lemma (Luger, 2003, p. 264) that no two errors obtain the same value with positive probability. Note that this is not implied by Assumption 2 in Luger (2003). An implication is that outcome variables with finite support are ruled out.

the final set of equations. The standard approach is to derive the distribution of the test statistic. Wright (2000) and Luger (2003) follow this approach, their data generating processes are very similar due to the assumption that the median error is 0. Our set of possible data generating processes is too rich so that one cannot hope to find a distribution-free test statistic. So how may one go about finding an exact test? Consider first model (a). Following Gossner and Schlag (2013), a natural approach could be to estimate b and then to bound the deviation of this estimate from its mean. However, an unbiased estimator of b does not exist as it belongs to a lagged variable. Moreover, we are not aware of finite sample bounds on the bias of estimators of b. Hence we could not advance in this direction, as successfully done by Gossner and Schlag (2013) for linear regressions with exogenous regressors.

Instead we adapt an approach similar to that used in the test of Dickey and Fuller (1979) and Luger (2003). Instead of trying to find out the true value of b, the idea is to act as if $b = b_0$ is fixed and to construct a test that is designed to uncover a violation of this presumption. Moving the lagged variable to the left hand side we set $W_t = Y_t - b_0 Y_{t-1}$ for $t \in \{1, .., T\}$ and then use the fact that $W_t = a + \varepsilon_t$ holds when H_0 is true.

Following Dufour (2003), we have to make some assumptions on the errors as exact nontrivial tests otherwise do not exist. Note that Gossner and Schlag (2013) instead put exogenous bounds on Y_t. We assume there exist some g_1 and g_2 known to the statistician such that $g_1 \leq \varepsilon_t \leq g_2$ holds almost surely for all t whenever H_0 is true. It is as if we are considering the joint null hypothesis H_0 : "$b = b_0$ and $g_1 \leq \varepsilon_t \leq g_2$". This generates confidence sets of the parameters b and $g_2 - g_1$. We also consider a framework in which the bounds g_1 and g_2 change over time, where future bounds are random and depend on past realizations. This approach offers the possibility to reject the original null hypothesis $H_0 : b = b_0$ and allows to derive confidence intervals for b.

To construct a test we need to be aware of what kind of data will be generated when the alterative hypothesis is true. This is because we wish to design a test that has power. Let b_1 denote the true value of b. If $b_1 < 1$ then the process approaches stationarity as t increases, EW_t converges to $\frac{a(1-b_0)}{1-b_1}$ over time. This convergence will depend on how far y_0 is from the unconditional mean, in particular the finite sample properties look very different if y_0 is close or far from $\frac{a(1-b_0)}{1-b_1}$. When $b_1 = 1$ then W_t is nonstationary, it is either a random walk without drift ($a = 0$) or one with drift ($a \neq 0$). Given these very different properties we may face when the alterative hypothesis is true we construct two separate tests for the cases $b_1 < 1$ and $b_1 = 1$. Later we show how to combine them into a single test.

Consider the case where $b_1 < 1$. To construct a test we need a property that holds under the null but not under the alternative. As the process is stationary both under the null and under the alternative we need to search for some other property. One may wish to proceed as in the Dickey-Fuller test, to utilize the fact that W_t does not depend on Y_{t-1}. Specifically one would like to regress W_t on Y_{t-1}, however we do not have an exact test for this as Y_{t-1} is not exogenous. It would of course be enough to investigate the relationship between W_t and Y_{t-1}, as under H_0 we have that $E\left(W_t Y_{t-1}\right) = aY_{t-1}$. It is not clear how to deal with the parameter a, ideally we would like to consider a relationship between Y_{t-1} and something that has mean 0. A simple approach is to subtract the mean of the future values of W from W_t, so consider $B_t = W_t - \frac{1}{T-t+1}\sum_{k=t+1}^{T} W_k$ and investigate the relationship between Y_{t-1} and B_t. In particular, we obtain $E\left(B_t Y_{t-1}\right) = 0$. We could then proceed by investigating $\sum_{t=1}^{T} B_t Y_{t-1}$, or more generally $\sum_{t=1}^{T} B_t f\left(Y_1, .., Y_{t-1}\right)$ for some function f, using the fact that this sum has mean 0 under H_0. Hoeffding (1963) provides an inequality for martingales with exogenous bounds on the range of the martingale differences. To apply this we would have to rewrite $\sum_{t=1}^{T} B_t Y_{t-1}$ as a martingale and to impose exogenous bounds on the range of each of the martingale differences. However, we would like to have a more flexible approach as the terms will depend on past observations. This is where the paper by van de Geer (2002) comes in, helping us in two respects. It shows us how to rewrite a sum as a sum of martingale differences and it contains an inequality that uses bounds that are measurable with respect to the earlier terms in the sum. To better use this approach we substantially improve the bound of van de Geer (2002). We then choose f in a way that we expect will lead to the highest rejection probability. Interestingly, the resulting test statistic no longer measures the correlation of earlier values of Y with later values of W. Instead, the fine-tuning of f leads to a test statistic that directly measures how well one can use $Y_1, .., Y_{t-1}$ in order to predict B_t. As B_t is linear in Y_k for $k \geq t$ we use $\hat{a}_t + \hat{b}_t Y_{k-1}$ to predict Y_k for $k \geq t$ where $\hat{a}_t$ and $\hat{b}_t$ are estimates of a and b using $Y_1, .., Y_{t-1}$. The test statistic is then compared to a cutoff that identifies how difficult it is to predict B_t under H_0. Note that it is very difficult to predict B_t under H_0 as B_t is uncorrelated with any random variable that is measurable with respect to what happened before t. As the true value b_1 of b moves away from the value b_0 of b under the null hypothesis, B_t starts to get correlated with its past which makes prediction easier. This leads to an increase in the rejection probability. However, when b_1 gets too close to 1 then we do very bad in predicting B_t based on $Y_1, .., Y_{t-1}$. Hence, when b_1 gets very close to 1 we have to consider an alternative method.

Consider the case where $b_0 < 1$ and $b_1 = 1$. Then W_t is stationary under H_0 while W_t is nonstationary under the alternative. For the model where the ranges of errors are constant we use the martingale inequality of Pinelis (2006), which is more powerful than that of Hoeffding (1963). As test statistic we take a weighted sum of the W's such that we have a martingale with mean 0 under H_0. If instead $b_1 = 1$ then it will have mean $\neq 0$ if $a \neq 0$, as W_t is nonstationary, the mean of the weighted sum of the W$'s$ can considerably differ from 0.

We now combine the two cases above. As we do not know the true value b_1 of b, we use the estimate $\hat{b}_t$ of b to determine which of the two tests above to use. We start with the case of prediction designed for $b_1 < 1$ and only turn to the case of $b_1 = 1$ if $\hat{b}_t$ is sufficiently close to 1. Once we switch to the test for $b_1 = 1$ we only use the remaining observations after the switch in order to guarantee that the level of the test remains bounded above by α.[5]

Using the same methodology we construct tests for model (ac). For this we need to change the definition of B_t to ensure that $E(B_t|F_{t-1}) = 0$. We do this by estimating a and c and proceeding similarly to model (a). For the test of $b_1 = 1$ we consider here long double differences $W_{T/2+1} - W_t$.

4 The Test

In future versions there will be a brief explanation of the testing procedure.

5 The Formal Details

We now derive our test in detail. As explained in Section 3, we design two separate tests, depending on whether they should generate rejections when the alternative is stationary ($b < 1$) or nonstationary ($b = 1$). For each of these two tests we first present a martingale inequality and then apply this to the autoregressive model. Later we combine them to a single test.

First we make some notes on the different inequalities. Hoeffding (1963, Theorem 2) derived an inequality for a sum of independent variables where the summands have exogenous bounds on their supports. He points out that it extends immediately to a sum of martingale differences. Pinelis (2006) provides an improvement of Hoeffding (1963, Theorem 2) for tail probabilities below 0.33. Van de Geer (2002) extends the inequality of Hoeffding (1963) to allow for ranges to depend on earlier observations. We improve this inequality to obtain a formula that is appropriate for our application. We use our improvement of van de Geer (2002) when targeting true data with $b < 1$ and the inequality of Pinelis (2006) when concerned with data with $b \geq 1$.

5.1 Tests for a Stationary Alternative

In this section, we design a test for the case where $b < 1$ under the alternative.

[5] If we would not do this then we would neglect the dependencies that arise by using them in determining the switch.

5.1.1 A Martingale Inequality and an Improvement

Recall the following inequality due to van de Geer (2002, Theorem 2.6). Let $\{F_t\}_{t=1}^{T}$ be a filtration, let Z be a F_T measurable random variable, let L_t and U_t be F_{t-1} measurable such that $L_t \leq E(Z|F_t) \leq U_t$ holds almost surely, $t = 1, .., T$. Let $C_T^2 = \sum_{t=1}^{T} (U_t - L_t)^2$.

Theorem 1 (van de Geer, 2002) *For $u > 0$ and $v > 0$,*

$$P\left(|Z - EZ| \geq u, C_T^2 \leq v^2\right) \leq 2\exp\left(-\frac{2u^2}{v^2}\right).$$

We improve this result, as follows directly when comparing the two statements.

Theorem 2 *For $u > 0$ and $v > 0$,*

$$P\left(|Z - EZ| \geq \frac{u}{2} + \frac{u}{2}\frac{C_T^2}{v^2}\right) \leq 2\exp\left(-\frac{2u^2}{v^2}\right).$$

Proof. Consider first the one-sided version. We follow closely the proof of Theorem 2.2 of van de Geer (2002).[6] Without loss of generality we can assume that $EZ = 0$. Let $X_t = E(Z|F_t) - E(Z|F_{t-1})$ for $t = 1, .., T$. Then $\{X_t\}_{t=1}^{T}$ is a martingale difference sequence and $Z = \sum_{t=1}^{T} X_t$. Consider some $\beta > 0$. Following (Hoeffding, 1963, 4.16), $E(\exp(\beta X_t)) \leq \exp\left(\beta^2 (U_t - L_t)^2 /8\right)$ almost surely (see also (van de Geer, 2002, Lemma 2.4)). So

$$E\left(\exp\left(\beta X_t - \beta^2 (U_t - L_t)^2 /8\right)\right) \leq 1.$$

Let $\zeta = \exp\left(\beta Z - \beta^2 C_T^2/8\right)$. Then

$$E\zeta = E\left(\prod_{t=1}^{T} \exp\left(\beta X_t - \beta^2 (U_t - L_t)^2 /8\right)\right) \leq 1.$$

Let $A = \{\zeta \geq e^r\}$. Note that $\zeta \geq e^r$ if A is true, hence $1 \geq E(\zeta \cdot 1_A) \geq \int_A e^r dP = e^r P(A)$ so $P(A) \leq e^{-r}$.

Now note that $\beta Z - \beta^2 C_T^2/8 \geq r$ if and only if $Z \geq \frac{r}{\beta} + \beta C_T^2/8$ so we set $\beta = \frac{4u}{v^2}$ and $r = \frac{2u^2}{v^2}$ to obtain

$$\frac{r}{\beta} + \beta C_T^2/8 = \frac{u}{2} + \frac{u}{2}\frac{C_T^2}{v^2}.$$

The two-sided inequality emerges when replacing Z with $-Z$ and then combining the two inequalities. ■

Next we consider probabilistic bounds defined as follows. At each time t there are two possible states l and h. Let s_t denote that state in period t, so $s_t \in \{l, h\}$, and

[6] The more general proof that proves that the inequality holds for some $Z_t = \sum_{i=1}^{t} X_i$ follows easily.

let η_t be the probability that $s_t = h$. There are F_{t-1} measurable random variables L_t^l, L_t^h, U_t^l and U_t^h such that $P\left(E\left(Z|F_t\right) \in \left[L_t^{\bar{s}_t}, U_t^{\bar{s}_t}\right] |s_t = \bar{s}_t\right) = 1$ for $\bar{s}_t \in \{l, h\}$ where $E\left(Z|F_t, s_t = \bar{s}_t\right)$ does not depend on $\bar{s}_t$. All other assumptions are now assumed to hold conditional on any sequence of states.

We will slightly update our inequality, as given in Theorem 2, now consider β_t that is F_{t-1} measurable.

Proposition 1 *Under the above assumptions, for any $\beta_t \geq 0$ that is F_{t-1} measurable,*

$$P\left(\begin{array}{c} \left|\sum_{t=1}^T \beta_t \left(E\left(Z|F_t\right) - E\left(Z|F_{t-1}\right)\right)\right| \\ -\sum_{t=1}^T \ln\left(\left(1-\eta_t\right)\exp\left(\beta_t^2\left(U_t^l - L_t^l\right)^2/8\right) + \eta_t \exp\left(\beta_t^2\left(U_t^h - L_t^h\right)^2/8\right)\right) \geq r \end{array}\right) \leq 2e^{-r}$$

Proof. As $E\left(X_i|s_i = \bar{s}_i, \bar{F}_{i-1}\right) = 0$ for $\bar{s}_i \in \{l, h\}$ we can separately apply (Hoeffding, 1963, (4.16)) to each state in period t. We namely compute

$$\begin{aligned} E\left(\exp\left(\beta_i X_i\right)|F_{i-1}\right) &= \left(1-\eta_i\right)E\left(\exp\left(\beta_i X_i\right)|F_{i-1}, s_i = l\right) + \eta_i E\left(\exp\left(\beta_i X_i\right)|F_{i-1}, s_i = h\right) \\ &\leq \left(1-\eta_i\right)\exp\left(\beta_i^2\left(U_i^l - L_i^l\right)^2/8\right) + \eta_i \exp\left(\beta_i^2\left(U_i^h - L_i^h\right)^2/8\right). \end{aligned}$$

Consequently,

$$\frac{E\left(\exp\left(\beta_i X_i\right)|F_{i-1}\right)}{\left(1-\eta_i\right)\exp\left(\beta_i^2\left(U_i^l - L_i^l\right)^2/8\right) + \eta_i \exp\left(\beta_i^2\left(U_i^h - L_i^h\right)^2/8\right)} \leq 1.$$

Let

$$\zeta = \frac{\exp\left(\sum_{t=1}^T \beta_t \left(E\left(Z|F_t\right) - E\left(Z|F_{t-1}\right)\right)\right)}{\prod_{i=1}^T \left(\left(1-\eta_i\right)\exp\left(\beta_i^2\left(U_i^l - L_i^l\right)^2/8\right) + \eta_i \exp\left(\beta_i^2\left(U_i^h - L_i^h\right)^2/8\right)\right)}.$$

Then $E\zeta \leq 1$. Continuing as in the proof of Theorem 2 we obtain the result. ■

5.1.2 Time Series Application

We now show how to use the above for our time series application. Fix $b_0 \in [0, 1]$ and consider $H_0 : b = b_0$. Let $W_t = Y_t - b_0 Y_{t-1}$ for $t \in \{1, .., T\}$. Let $\bar{F}$ be the filtration defined by $\{W_t\}$.

We apply our inequality to $Z = \sum_{t=1}^T \hat{S}_t B_t$ where $B_t = W_t - \sum_{k=t}^T \tau_{tk} W_k$ and $\hat{S}_t$ is some $\bar{F}_{t-1}$ measurable random variable. Setting $A_i = \hat{S}_i - \sum_{t=1}^i \tau_{ti}\hat{S}_t$, so A_i is $\bar{F}_{i-1}$ measurable, we obtain that $Z = \sum_{i=1}^T A_i W_i$. We choose $\{\tau_{tk}\}$ such that $E\left(B_t|\bar{F}_{t-1}\right) = 0$ under H_0. This means that $\sum_{k=t}^T \tau_{tk} W_k$ is an unbiased estimate of W_t. As a natural choice we set $\sum_{k=t}^T \tau_{tk} W_k$ equal to the OLS estimate of W_t based on $\{W_k\}_{k \geq t}$. For model (a), as $EW_t = a$ under H_0, we obtain $\tau_{tk} = \frac{1}{T-t+1}$ for $t \leq k \leq T$.

In particular, $\tau_{TT}=1$ and hence $B_T=0$. For model (ac) we have $EW_t=a+ct$, straight forward calculations then show that

$$\tau_{tk}=2\frac{t+1+2T}{(T-t+2)(T-t+1)}-6\frac{k}{(T-t+2)(T-t+1)}$$

for $t\leq k\leq T-2$ and $\tau_{tt}=1$ for $t\in\{T-1,T\}$ and $\tau_{T-1,T}=0$. In particular, $B_t=0$ for $t\in\{T-1,T\}$.

To simplify notation, set $T_M=T-1$ in model (a) and $T_M=T-2$ in model (ac). Given the above, w.l.o.g. we can set $\hat{S}_t=0$ for $t>T_M$. An additional property of the OLS estimates we use later that results is that $\sum_{i=k}^{T}\tau_{ki}\tau_{ti}=\tau_{tk}$ holds for $t\leq k\leq T_M$. This is immediate in model (a). For model (ac) it follows as for each t there exist constants d_{t1} and d_{t2} such that $\tau_{ti}=d_{t1}+d_{t2}\cdot i$, hence τ_{ti} has the same dependency on t as W_t under H_0. Formally, we verify

$$\sum_{i=k}^{T}\tau_{ki}\tau_{ti}=\sum_{i=k}^{T}\tau_{ki}\left(d_{t1}+d_{t2}\cdot i\right)=d_{t1}+d_{t2}\cdot k=\tau_{tk}\text{ for }t\leq k.$$

The next step is to derive $E\left(Z|\bar{F}_i\right)-E\left(Z|\bar{F}_{i-1}\right)$ under H_0. For $t\leq k$,

$$E\left(\hat{S}_tW_k|\bar{F}_i\right)-E\left(\hat{S}_tW_k|\bar{F}_{i-1}\right)=\begin{cases}E\left(\hat{S}_tW_k|\bar{F}_i\right)-E\left(\hat{S}_tW_k|\bar{F}_{i-1}\right) & \text{for }i+1\leq t\leq k\\ 0 & \text{for }t\leq i<k\\ \hat{S}_t\left(W_i-E\left(W_i|\bar{F}_{i-1}\right)\right) & \text{for }t\leq k=i\\ 0 & \text{for }t\leq k\leq i-1\end{cases}.$$

So

$$\begin{aligned}&\sum_{t=1}^{T}E\left(\hat{S}_tW_t|\bar{F}_i\right)-\sum_{t=1}^{T}E\left(\hat{S}_tW_t|\bar{F}_{i-1}\right)\\ =\;&\hat{S}_i\left(W_i-E\left(W_i|\bar{F}_{i-1}\right)\right)+\sum_{t=i+1}^{T_M}E\left(\hat{S}_tW_t|\bar{F}_i\right)-E\left(\hat{S}_tW_t|\bar{F}_{i-1}\right)\\ =\;&\hat{S}_i\left(W_i-E\left(W_i|\bar{F}_{i-1}\right)\right)\end{aligned}$$

and

$$\begin{aligned}&\sum_{t=1}^{T}\left(E\left(\hat{S}_t\sum_{k=t}^{T}\tau_{tk}W_k|\bar{F}_i\right)-E\left(\hat{S}_t\sum_{k=t}^{T}\tau_{tk}W_k|\bar{F}_{i-1}\right)\right)\\ =\;&\sum_{t=1}^{T}\sum_{k=t}^{T}\tau_{tk}\left(E\left(\hat{S}_tW_k|\bar{F}_i\right)-E\left(\hat{S}_tW_k|\bar{F}_{i-1}\right)\right)\\ =\;&\sum_{t=1}^{i-1}\sum_{k=t}^{i-1}...+\sum_{t=1}^{i}\sum_{k=i}^{i}...+\sum_{t=1}^{i}\sum_{k=i+1}^{T}...+\sum_{t=i+1}^{T}\sum_{k=t}^{T}...\\ =\;&0+\sum_{t=1}^{i}\tau_{ti}\hat{S}_t\left(W_i-E\left(W_i|\bar{F}_{i-1}\right)\right)+0+\sum_{t=i+1}^{T}\sum_{k=t}^{T}\tau_{tk}\left(E\left(\hat{S}_tW_k|\bar{F}_i\right)-E\left(\hat{S}_tW_k|\bar{F}_{i-1}\right)\right)\end{aligned}$$

$$\begin{aligned}
&= \sum_{t=1}^{i} \tau_{ti} \hat{S}_t \left(W_i - E\left(W_i | \bar{F}_{i-1} \right) \right) + \sum_{t=i+1}^{T} \left(E\left(\hat{S}_t \sum_{k=t}^{T} \tau_{tk} W_k | \bar{F}_i \right) - E\left(\hat{S}_t \sum_{k=t}^{T} \tau_{tk} W_k | \bar{F}_{i-1} \right) \right) \\
&= \sum_{t=1}^{i} \tau_{ti} \hat{S}_t \left(W_i - E\left(W_i | \bar{F}_{i-1} \right) \right) + \sum_{t=i+1}^{T} \left(E\left(\hat{S}_t W_t | \bar{F}_i \right) - E\left(\hat{S}_t W_t | \bar{F}_{i-1} \right) \right) \\
&= \sum_{t=1}^{i} \tau_{ti} \hat{S}_t \left(W_i - E\left(W_i | \bar{F}_{i-1} \right) \right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
E\left(Z | \bar{F}_i \right) - E\left(Z | \bar{F}_{i-1} \right) &= \left(\hat{S}_i - \sum_{t=1}^{i} \tau_{ti} \hat{S}_t \right) \left(W_i - E\left(W_i | \bar{F}_{i-1} \right) \right) \\
&= A_i \left(W_i - E\left(W_i | \bar{F}_{i-1} \right) \right)
\end{aligned}$$

and hence,

$$E\left(Z | \bar{F}_i \right) = E\left(Z | \bar{F}_{i-1} \right) - A_i E\left(W_i | \bar{F}_{i-1} \right) + A_i W_i.$$

Under H_0 we have that $W_k = a + c \cdot k + \varepsilon_k$ with $g_{k1} \leq \varepsilon_k \leq g_{k2}$ where $c = 0$ if we are considering model (a). So let $C_k = E\left(Z | \bar{F}_{k-1} \right) - A_k E\left(W_k | \bar{F}_{k-1} \right) + A_k \cdot (a + c \cdot k)$, $L_k = C_k + \min\{A_k g_{k1}, A_k g_{k2}\}$ and $U_k = C_k + \max\{A_k g_{k1}, A_k g_{k2}\}$. Then L_k and U_k are $\bar{F}_{k-1}$ measurable and $E\left(Z | \bar{F}_k \right) \in [L_k, U_k]$ almost surely. Let $C_T^2 = 4 \sum_{t=1}^{T} g_t^2 A_t^2$ and $r = \frac{2u^2}{v^2}$. Since $E(Z) = 0$ by construction we then obtain from Theorem 2 that

$$\begin{aligned}
&P\left(\sum_{t=1}^{T} \hat{S}_t B_t \geq \frac{u}{2} + \frac{r}{u} \sum_{t=1}^{T} g_t^2 A_t^2 \right) \\
&= P\left(\sum_{t=1}^{T} A_t \left(W_t - \frac{r}{u} g_t^2 A_t \right) \geq \frac{u}{2} \right) \leq e^{-r}. \qquad (1)
\end{aligned}$$

Given this expression one might choose to maximize $\sum_{t=1}^{T} A_t \left(W_t - \frac{r}{u} g_t^2 A_t \right)$ over A_t. A first problem is that A_t cannot be chosen, only $\hat{S}_t$ can be chosen. For $k \leq T_M$ one can solve A_k for $\hat{S}_k$ and act as if A_k can be chosen. However this has impact on A_k for $k > T_M$ that we cannot directly control as A_k does not depend on $\hat{S}_k$ for $k > T_M$. A second problem is that g_t is not known at time k for $t > k$. This second problem we will deal with by assuming when deriving $\hat{S}_k$ at time k that all future $g's$ are expected to be equal, so when choosing $\hat{S}_k$ we will act as if $g_t = g_k$ for all $t > k$.

To circumvent the first problem we will rewrite this expression as a function of $\hat{S}$. We will show for each k that there is some function f_k such that

$$\sum_{t=1}^{T} g_t^2 A_t^2 = f_k\left(\hat{S}_1, .., \hat{S}_{k-1} \right) + g_k^2 \sum_{t=k}^{T_M} (1 - \tau_{tt}) \hat{S}_t^2 \text{ if } g_t = g_k \text{ for } t > k.$$

In particular, if $g_k = g$ does not depend on k we obtain

$$g^2 \sum_{t=1}^{T} A_t^2 = g^2 \sum_{t=1}^{T_M} (1 - \tau_{tt}) \hat{S}_t^2.$$

We prove this as follows. For $k \le T_M$ consider

$$\frac{d}{d\hat{S}_k}\left(\frac{1}{2}\sum_{t=1}^{T} g_t^2 A_t^2\right) = g_k^2 \hat{S}_k - g_k^2 \sum_{t=1}^{k} \tau_{tk}\hat{S}_t - \sum_{i=k}^{T} \tau_{ki} g_i^2 \left(\hat{S}_i - \sum_{t=1}^{i} \tau_{ti}\hat{S}_t\right)$$

$$= g_k^2 \hat{S}_k - g_k^2 \sum_{t=1}^{k} \tau_{tk}\hat{S}_t - \sum_{i=k}^{T} g_i^2 \tau_{ki}\hat{S}_i + \sum_{i=k}^{T}\sum_{t=1}^{i} g_i^2 \tau_{ki}\tau_{ti}\hat{S}_t$$

$$= g_k^2 \hat{S}_k - g_k^2 \sum_{t=1}^{k} \tau_{tk}\hat{S}_t - \sum_{i=k}^{T} g_i^2 \tau_{ki}\hat{S}_i + \sum_{i=k}^{T} g_i^2 \sum_{t=1}^{k-1} \tau_{ki}\tau_{ti}\hat{S}_t + \sum_{i=k}^{T} g_i^2 \sum_{t=k}^{i} \tau_{ki}\tau_{ti}\hat{S}_t$$

$$= g_k^2 \hat{S}_k - g_k^2 \sum_{t=1}^{k} \tau_{tk}\hat{S}_t - \sum_{i=k}^{T} g_i^2 \tau_{ki}\hat{S}_i + g_k^2 \sum_{t=1}^{k-1}\left(\sum_{i=k}^{T} \tau_{ki}\tau_{ti}\right)\hat{S}_t + g_k^2 \sum_{t=k}^{T}\left(\sum_{i=t}^{T} \tau_{ki}\tau_{ti}\right)\hat{S}_t$$

where we used the assumption that $g_i = g_k$ for all $i > k$. Using the fact τ satisfies $\sum_{i=k}^{T} \tau_{ki}\tau_{ti} = \tau_{tk}$ for $t \le k \le T_M$ we obtain

$$\frac{d}{d\hat{S}_k}\left(\frac{1}{2}\sum_{t=1}^{T} g_t^2 A_t^2\right) = g_k^2 \hat{S}_k - g_k^2 \sum_{t=1}^{k} \tau_{tk}\hat{S}_t - g_k^2 \sum_{i=k}^{T} \tau_{ki}\hat{S}_i + g_k^2 \sum_{t=1}^{k-1} \tau_{tk}\hat{S}_t + g_k^2 \sum_{t=k}^{T} \tau_{kt}\hat{S}_t = g_k^2\left(1-\tau_{kk}\right)\hat{S}_k.$$

This together with the linearity of $\sum_{t=1}^{T} g_t^2 A_t^2$ in $\hat{S}_k$ proves the statement.

Hence the test statistic at time k is believed to be

$$\phi_k(u) := \sum_{t=1}^{T} \hat{S}_t B_t - \frac{r}{u}\sum_{t=1}^{T} g_t^2 A_t^2 = \sum_{t=1}^{k-1} \hat{S}_t B_t - \frac{r}{u} f_k + \sum_{t=k}^{T} \hat{S}_t \left(B_t - \left(1-\tau_{tt}\right)\frac{r}{u} g_k^2 \hat{S}_t\right)$$

$$= \sum_{t=1}^{k-1} \hat{S}_t B_t - \frac{r}{u} f_k + \sum_{t=k}^{T_M} \hat{S}_t \left(B_t - \left(1-\tau_{tt}\right)\frac{r}{u} g_k^2 \hat{S}_t\right).$$

We now consider the choice of $\hat{S}_k$ for $k \le T_M$. Ideally we would set the derivative of ϕ equal to 0, but as we do not know B_t at time t we replace B_t by $E\left(B_t|\bar{F}_{t-1}\right)$ and then set

$$\hat{S}_k = \frac{u}{2\left(1-\tau_{kk}\right) r g_k^2} E\left(B_k|\bar{F}_{k-1}\right) \tag{2}$$

to obtain

$$\phi_k(u) = \sum_{t=1}^{k-1} \hat{S}_t B_t - \frac{r}{u} f_k + \frac{u}{2rg_k^2}\frac{1}{1-\tau_{kk}} E\left(B_k|\bar{F}_{k-1}\right)\left(B_k - \frac{1}{2}E\left(B_k|\bar{F}_{k-1}\right)\right)$$

$$+ \sum_{t=k+1}^{T_M} \hat{S}_t \left(B_t - \left(1-\tau_{tt}\right)\frac{r}{u} g_k^2 \hat{S}_t\right).$$

We now continue in the special case where $g_t = g$ does not depend on t. This yields a simple formula. Remember that $P\left(\phi(u) \ge \frac{u}{2}\right) \le e^{-r}$. Whether or not $\phi(u) \ge u/2$ does not depend on the value of $u > 0$ and we obtain the following result.

Proposition 2 *If H_0 is true and $g_t = g$ for all t then*

$$P\left(\frac{1}{g^2}\sum_{t=1}^{T_M}\frac{1}{1-\tau_{tt}}E\left(B_t|\bar{F}_{t-1}\right)\left(B_t-\frac{1}{2}E\left(B_t|\bar{F}_{t-1}\right)\right)\geq r\right)\leq e^{-r}.$$

Entering the expression for τ_{tt} then yields the formulae provided in Section 4. Finally note that the expression for $E\left(W_k|\bar{F}_{t-1}\right)$ for $k \geq t$ given in Section 4 is easily derived from the model.

Consider now the more general case where g_t may be F_{t-1} measurable. We insert the values of $\hat{S}_t$ for $t \leq T_M$ from (2), maintaining $\hat{S}_t = 0$ for $t > T_M$, to obtain

$$\begin{aligned} A_t &= \hat{S}_t - \sum_{k=1}^{t}\tau_{kt}\hat{S}_k = \frac{u}{2\left(1-\tau_{tt}\right)rg_t^2}E\left(B_t|\bar{F}_{t-1}\right) - \sum_{k=1}^{t}\tau_{kt}\frac{u}{2\left(1-\tau_{kk}\right)rg_k^2}E\left(B_k|\bar{F}_{k-1}\right) \text{ for } t\leq T_M \\ A_t &= -\sum_{k=1}^{T_M}\tau_{kt}\frac{u}{2\left(1-\tau_{kk}\right)rg_k^2}E\left(B_k|\bar{F}_{k-1}\right) \text{ for } t>T_M \end{aligned}$$

which we then insert into (1).

Proposition 3 *If H_0 is true then for g_t that is F_{t-1} measurable for $t = 1,..,T$ we obtain*

$$P\left(\begin{array}{c} \sum_{t=1}^{T_M}\frac{1}{(1-\tau_{tt})g_t^2}E\left(B_t|\bar{F}_{t-1}\right)B_t \\ -\frac{1}{2}\sum_{t=1}^{T_M}\left(\frac{1}{(1-\tau_{tt})g_t}E\left(B_t|\bar{F}_{t-1}\right)-\sum_{k=1}^{t}\frac{\tau_{kt}g_t}{(1-\tau_{kk})g_k^2}E\left(B_k|\bar{F}_{k-1}\right)\right)^2 \\ +\frac{1}{2}\sum_{t=T_M+1}^{T}\left(\sum_{k=1}^{T_M}\frac{\tau_{kt}g_t}{(1-\tau_{kk})g_k^2}E\left(B_k|\bar{F}_{k-1}\right)\right)^2 \geq r \end{array}\right)\leq e^{-r}.$$

Next we consider probabilistic bounds. We make the following assumptions. At each time t there are two possible states l and h. Let s_t denote that state in period t, so $s_t \in \{l,h\}$, and let η_t be the probability that $s_t = h$. Moreover, there are F_{t-1} measurable random variables g_{t1}^l, g_{t1}^h, g_{t2}^l and g_{t2}^h such that $g_{t1}^{\bar{s}_t} < g_{t2}^{\bar{s}_t}$ and $\varepsilon_t|_{s_i=\bar{s}_i} \in \left[g_{t1}^{\bar{s}_t}, g_{t2}^{\bar{s}_t}\right]$ for $\bar{s}_t \in \{l,h\}$. Let $g_t^{\bar{s}_t} = \frac{1}{2}\left(g_{t2}^{\bar{s}_t} - g_{t1}^{\bar{s}_t}\right)$ for $\bar{s}_t \in \{l,h\}$.

We will slightly update our inequality. Specifically we will utilize the fact that

$$\begin{aligned} X_i|_{s_i=\bar{s}_i} &= E\left(Z|\bar{F}_i\right)|_{s_i=\bar{s}_i} - E\left(Z|\bar{F}_{i-1}\right)|_{s_i=\bar{s}_i} = A_i\left(W_i|_{s_i=\bar{s}_i} - E\left(W_i|\bar{F}_{i-1}\right)\right) \\ &= A_i\left(a+ci+\varepsilon_i|_{s_i=\bar{s}_i}\right) - A_i\left(a+ci\right) = A_i\varepsilon_i|_{s_i=\bar{s}_i} \end{aligned}$$

and hence $E\left(X_i|s_i=\bar{s}_i, \bar{F}_{i-1}\right) = 0$ for $\bar{s}_i \in \{l,h\}$. We can thus apply Proposition 1. With $\left(U_t^{\bar{s}_t} - L_t^{\bar{s}_t}\right)^2 = 4\left(g_t^{\bar{s}_t}\right)^2 A_t^2$, and setting $\beta = \frac{2r}{u}$ we obtain

$$P\left(\frac{2r}{u}\sum_{t=1}^{T_M}\hat{S}_tB_t - \sum_{t=1}^{T}\ln\left(\left(1-\eta_t\right)\exp\left(\frac{2r^2}{u^2}\left(g_t^l\right)^2A_t^2\right)+\eta_t\exp\left(\frac{2r^2}{u^2}\left(g_t^h\right)^2A_t^2\right)\right)\geq r\right)\leq e^{-r}.$$

Setting $\bar{g}_t^2 = (1-\eta_t)\left(g_t^l\right)^2 + \eta_t\left(g_t^h\right)^2$ we then choose

$$\hat{S}_t = \frac{u}{2\left(1-\tau_{tt}\right) r\bar{g}_t^2} E\left(B_t|\bar{F}_{t-1}\right),$$

choose H_t such that $A_t = \frac{u}{2r} H_t$ where H_t is independent of u and r, so

$$\begin{aligned} H_t &= \frac{1}{(1-\tau_{tt})}\frac{E\left(B_t|\bar{F}_{t-1}\right)}{\bar{g}_t^2} - \sum_{k=1}^{t}\tau_{kt}\frac{1}{1-\tau_{kk}}\frac{E\left(B_k|\bar{F}_{k-1}\right)}{\bar{g}_k^2} \text{ for } t \leq T_M \\ H_t &= -\sum_{k=1}^{T_M}\tau_{kt}\frac{1}{1-\tau_{kk}}\frac{E\left(B_k|\bar{F}_{k-1}\right)}{\bar{g}_k^2} \text{ for } t > T_M. \end{aligned}$$

This leads to the following result.

Proposition 4 *If H_0 is true then under the above model with probabilistic ranges, then*

$$P\left(\begin{array}{c} \sum_{t=1}^{T_M}\frac{1}{(1-\tau_{tt})\bar{g}_t^2}E\left(B_t|\bar{F}_{t-1}\right)B_t \\ -\sum_{t=1}^{T}\ln\left((1-\eta_t)\exp\left(\frac{1}{2}H_t^2\cdot\left(g_t^l\right)^2\right)+\eta_t\exp\left(\frac{1}{2}H_t^2\cdot\left(g_t^h\right)^2\right)\right) \geq r \end{array}\right) \leq e^{-r}.$$

5.2 Tests for a Nonstationary Alternative

Next, we design a test for the case where $b = 1$ (or is approximately equal to 1) under the alternative hypothesis.

5.2.1 A Martingale Inequality

We recall an inequality of Pinelis (2006, Corollary 2.2, Theorem 3.1). Let $\{Z_t\}_{t=1}^{T}$ be be a martingale difference sequence with respect to the filtration $\{F_t\}$. Assume that $Z_i - EZ_i \in [C_{i-1}, D_{i-1}]$ almost surely with $D_{i-1} - C_{i-1} \leq d_i$ and $d^2 = \sum_{i=1}^{T} d_i^2$ where d_i is real and C_{i-1}, D_{i-1} are F_{i-1} measurable. The two-sided version is given by

$$P\left(\left|\sum_{i=1}^{T}(Z_i - EZ_i)\right| \geq rd/2\right) \leq 2\min\left\{5!\left(\frac{e}{5}\right)^5(1-\Phi(r)), \exp\left(-\frac{r^2}{2}\right)\right\}$$

where $5!\left(\frac{e}{5}\right)^5 = 5.699...$ and Φ is the cdf of the standard normal distribution. The bound $\exp\left(-\frac{r^2}{2}\right)$ is that of Hoeffding (1963, Theorem 2) generalized to the situation where the width of the range rather than the range itself is exogenous. Note that the part of the bound using Φ is binding, which makes this inequality an improvement over that of Hoeffding (1963), when the right hand side is below 0.336.

5.2.2 Time Series Application for Deterministic Bounds

To be able to apply the above we need that the ranges have known bounds. Consider first the case where g_t is nonrandom for each t. For expositional purposes, and as this is the main case of interest, we additionally assume that $g_t = g$ is independent of t.

Consider model (a) where $W_t = a + \varepsilon_t$ when H_0 is true. Consider some $M \in \{1, .., T-1\}$ that has to be chosen before observing the data. Let $Z_k = W_k - a$ for $k \leq M$ and $Z_k = \frac{M}{T-M}(a - W_k)$ for $k = M+1, .., T$. Then Z_t is a martingale with respect to the filtration $\left(\bar{F}_t\right)_t$ generated by $\{W_t\}$. Moreover, $Z_k \in [g_1, g_2]$ for $k \leq M$, $Z_k \in \left[-\frac{M}{T-M}g_2, -\frac{M}{T-M}g_1\right]$ for $k > M$ and $E\left(\sum_{k=1}^T Z_k\right) = 0$. Setting $C_k = g_1$, $D_k = g_2$ and $d_k = 2g$ for $k \leq M$ and $C_k = -\frac{M}{T-M}g_2$, $D_k = -\frac{M}{T-M}g_1$ and $d_k = 2\frac{M}{T-M}g$ for $k > M$ we apply the inequality of Pinelis, with $d^2 = M4g^2 + (T-M)\,4\left(\frac{M}{T-M}\right)^2 g^2 = 4g^2\frac{TM}{T-M}$.

Proposition 5 *For model (a), assume that $g_t = g$ for each t. If H_0 is true then*

$$\begin{aligned} & P\left(\left|\frac{1}{M}\sum_{t=1}^{M} W_t - \frac{1}{T-M}\sum_{t=M+1}^{T} W_t\right| \geq rg\sqrt{\frac{T}{M\,(T-M)}}\right) \\ \leq\; & 2\min\left\{5!\left(\frac{e}{5}\right)^5 (1-\Phi(r)), \exp\left(-\frac{r^2}{2}\right)\right\}. \end{aligned} \qquad (3)$$

Now consider model (ac) where $W_t = a + ct + \varepsilon_t$ when H_0 is true. Here adapt a more general approach and consider as test statistic $\sum_{t=1}^T \gamma_t W_t$ where (γ_t) are constants such that $\sum_{t=1}^T \gamma_t = 0$ and $\sum_{t=1}^T t\gamma_t = 0$. It then follows that $E\left(\sum_{t=1}^T \gamma_t W_t\right) = 0$ holds when H_0 is true. So $d_i = 2g\,|\gamma_t|$ and $d^2 = 4g^2 \sum_{t=1}^T \gamma_t^2$.

Proposition 6 *For model (ac), assume that g_t is nonrandom for each t. If H_0 is true then*

$$P\left(\left|\sum_{t=1}^T \gamma_t W_t\right| \geq rg\sqrt{\sum_{t=1}^T \gamma_t^2}\right) \leq 2\min\left\{5!\left(\frac{e}{5}\right)^5 (1-\Phi(r)), \exp\left(-\frac{r^2}{2}\right)\right\}.$$

In the simulations, we choose γ in the following symmetric way: Given $M \leq T/2$ set $\gamma_t = 1$ for $t \leq M$ and $t \geq T - M + 1$ and $\gamma_t = -\frac{2M}{T-2M}$ for $M < t \leq T - M$. In particular, $M = T/4$ and hence $|\gamma_t| = 1$ for all t.

5.2.3 Time Series Application for Probabilistic Bounds

Consider now the model with probabilitistic bounds as defined at the end of Section 5.1.2. Here we have to revert to our improvement of the inequality of van de Geer (2002).

Consider model (a). Consider β_t that are F_{t-1} measurable such that $\sum_{t=1}^{T}\beta_t = 0$ almost surely. Then $E\left(\sum_{t=1}^{T}\beta_t W_t\right) = 0$. We apply Proposition 1 and obtain

$$P\left(\begin{array}{c}\left|\sum_{t=1}^{T}\beta_t W_t\right| \geq \\ \sum_{t=1}^{T}\ln\left((1-\eta_t)\exp\left(\beta_t^2\left(g_t^l\right)^2/2\right)+\eta_t\exp\left(\beta_t^2\left(g_t^h\right)^2/2\right)\right)+r\end{array}\right)\leq 2e^{-r}.$$

We are only left with choosing the values of β.

One way to proceed is as follows. Consider some $M \in \mathbb{N}$. To simplify presentation, assume $T \bmod M = 0$. The first $\frac{M-1}{M}T$ values of β are chosen freely subject to being measurable with respect to the past. The last T/M value of β are then the negative values of these first ones. Let $k = l(M-1)+i$ with $k \leq \frac{M-1}{M}T$, $l, i \in \mathbb{N}_0$ with $i \leq M-1$ (so $i = 1+(k-1) \bmod (M-1)$ and $l = \lfloor (k-1)/(M-1) \rfloor$). So there are T/M blocks of size $M-1$ in the first $\frac{M-1}{M}T$ observations, all indexes in the $(l+1)$st block are paired with the index $\frac{M-1}{M}T+l+1$. Accordingly, we choose β_k to be F_{k-1} measurable for $k \leq \frac{M-1}{M}T$ and we then set $\beta_{\frac{M-1}{M}T+l+1} = -\sum_{t=l(M-1)+1}^{(l+1)(M-1)}\beta_t$. This ensures $\sum_{t=1}^{T}\beta_t = 0$.

Consider first the case where g^l, g^h and η do not depend on t. We choose β independent of t and obtain

$$P\left(\begin{array}{c}\sum_{k=1}^{\frac{M-1}{M}T}W_k-(M-1)\sum_{l=0}^{T/M-1}W_{\frac{M-1}{M}T+l+1} \geq \\ \frac{T}{\beta}\ln\left((1-\eta)\exp\left(\beta^2\left(g^l\right)^2/2\right)+\eta\exp\left(\beta^2\left(g^h\right)^2/2\right)\right)+\frac{r}{\beta}\end{array}\right)\leq e^{-r}.$$

Ideally we would choose β to minimize

$$\frac{T}{\beta}\ln\left((1-\eta)\exp\left(\beta^2\left(g^l\right)^2/2\right)+\eta\exp\left(\beta^2\left(g^h\right)^2/2\right)\right)+\frac{r}{\beta}.$$

When η is small we minimize the analytically more tractable expression

$$\beta T\left((1-\eta)\left(g^l\right)^2+\eta\left(g^h\right)^2\right)/2+\frac{r}{\beta}.$$

The solution is given by

$$\bar{\beta} = \sqrt{\frac{2r}{T\left((1-\eta)\left(g^l\right)^2+\eta\left(g^h\right)^2\right)}}. \tag{4}$$

Thus we obtain

$$P\left(\begin{array}{c}\bar{\beta}\left|\sum_{k=1}^{\frac{M-1}{M}T}W_k-(M-1)\sum_{l=0}^{T/M-1}W_{\frac{M-1}{M}T+l+1}\right| \\ -T\ln\left((1-\eta)\exp\left(\bar{\beta}^2\left(g^l\right)^2/2\right)+\eta\exp\left(\bar{\beta}^2\left(g^h\right)^2/2\right)\right)\geq r\end{array}\right)\leq 2e^{-r}.$$

One can also emulate this solution when ranges and state probabilities are time dependent, acting in each period as if all periods had the values of the given period. The suggestion is to choose

$$\beta_k = \sqrt{\frac{2r}{T\left((1-\eta_k)\left(g_k^l\right)^2+\eta_k\left(g_k^h\right)^2\right)}} \tag{5}$$

for $k \leq \frac{M-1}{M}T$ and to choose

$$\beta_{\frac{M-1}{M}T+l+1} = -\sum_{t=l(M-1)+1}^{(l+1)(M-1)} \beta_t$$

for $l = 0, .., T/M - 1$.

In practical examples this choice has shown to have good performance.

Consider now model (ac). Here we have to choose β such that $\sum_{t=1}^{T} \beta_t = 0$ and $\sum_{t=1}^{T} t\beta_t = 0$. Assume for simplicity that $T \bmod M = 0$.

We again choose to proceed symmetrically, let $M = 3$, for $k \leq T/3$ choose β_k such that F_{k-1} measurable and let $\beta_{T/3+k} = -2\beta_k$ and $\beta_{2T/3+k} = \beta_k$ as $1 - 2 + 1 = 0$ and $k - 2\,(T/3 + k) + (2T/3 + k) = 0$.

To determine β we focus again on the case where g and η do not depend on t and choose β independent of t. The calculations are the same as in model (a), we obtain β defined by (4). For the case where g and η are adaptive we propose to use the formula for β_k given in (5).

5.3 Implementing the tests

The key to our analysis is that errors have given bounds. The existence of these bounds enables us to derive inequalities that are used to reject the null hypothesis. The rejection occurs when the data is sufficiently atypical from the perspective of the null hypothesis. The bounds themselves may also be violated by the data in which case one can also reject the null hypothesis. Whether or not the bounds are violated will depend on the null hypothesis as the errors themselves are not observable. Evidence of a violation of the bounds, hence a rejection, is given when they are incompatible with each choice of the parameters that is consistent with the null hypothesis. These two parts of the test, rejection based on inequality and rejection based on violation of the bounds are independent of each other. We put them into one test by splitting up the nominal level α. Let $\alpha_I \in (0, \alpha)$ be allocated to the test based on the inequalities and α_B be assigned to the bounds. Independence of these two tests implies that $1 - (1 - \alpha_I)\,(1 - \alpha_B) = \alpha$, so $\alpha_B = \frac{\alpha - \alpha_I}{1 - \alpha_I}$. Eg, if $\alpha = 0.05$, $\alpha_I = 0.025$ then $\alpha_B \approx 0.02564$.

So where do the bounds come from? One approach is to include the bounds in the null hypothesis, an alternative approach will be explained later. It is easy and insightful to include bounds in the null hypothesis when bounds do not change over time. One obtains for each value of the bound whether or not a specific value of the autoregressive coefficient can be rejected. This generates joint confidence sets for the value of the bounds and for the value of the autoregressive coefficient. This approach however has two downsides. One cannot reject any specific value of the autoregressive coefficient when the bounds are chosen to be sufficiently large. This is due to the fact that nontrivial tests do not exist when these bounds are not present. So the first downside is that the applied statistician can only reject a specific value of the autoregressive coefficient if he or she specifies the bound before seeing the data. The second downside is that joint confidence sets that include the values of the bounds is not very useful if one is interested in bounds that change over time. The joint confidence set for this case would be too rich as it would qualify the set of all time changing bounds under which the specific value of the autoregressive coefficient cannot be rejected. In view of our original motive to model heteroskedastic errors it is however natural that bounds change over time. So we include this approach with constant bounds as a benchmark, for comparisons between different data sets and for the case where the statistician can credibly commit to a specific value of the bounds before gathering the data.

Bounds are modelled probabilistic in order to formally allow for outliers. Outliers should not be eliminated adhoc, instead we include themc as follows. At each point in time errors are either drawn from a range that has a small width $2g^l$ or a large width $2g^h$. The realization of the width is assumed to be independent over time and independent of any other events. The probability that the range is large, so equal to $2g^h$, is constant and denoted by η where η is small. For illustration of the results we fix the ratio ρ between the small and large width, so $\rho = g^h/g^l$, and illustrate the joint confidence interval of b and g^l. Note that only the width of the interval enters as the inequalities only depend on this, and not on the location of the interval.

Incompatability with these probabilistic bounds is determined as follows. The large bounds are violated if there is some period t such that no parameters result in errors that lie within the bounds of this period. The small bounds are violated if sufficiently many errors necessarily lie outside the small bounds. Violation of the small bounds can be tested with the binomial test as the ranges are assumed to be realized independently of each other. Whether too many lie outside is checked by fixing the values of a and c and applying the binomial test to all outliers. If for each value of a and c the binomial test leads to a rejection then this part of the test leads to a rejection.

We now introduce an alternative approach in which ranges change over time. The way in which ranges change is embedded in the model, inference now focusses alone on whether or not a specific value b of the autoregressive coefficient can be rejected. Remember that the time series model is not an objective description of the data generating process. It is more a thought exersize according to which one wishes to evaluate the data. As described above, rejections can also be triggered based on the ranges. In the following we will describe how these ranges are determined. The mathematics shows that one can choose the range for the error at time t as function of the data that occurred up to time $t-1$. So the statistician thinks right before time t about where the errors in time t will lie. Errors depend on the unknown parameters, in this case a, b and c (note that c only plays a role in model (ac)). The statistician does not think about the null hypothesis in which the value b_0 of b is fixed but instead generally thinks about her understanding. So she estimates with her knowledge up to time $t-1$ the values of these parameters and uses these to get an understanding of the errors that occurred in the last K rounds. As any finite sample of errors will not cover the entire range, the range of these errors is inflated around the expected value in time t by a factor $\varphi > 1$. This determines the small range that occurs most of the time (with probability $1-\eta$ where η is small). But what happens to the estimates or understanding if the errors realized in period t, evaluated with the knowledge up to time $t-1$ fall outside the anticipated small range. When this occurs once the statistician believes that it is an outlier, that it belongs to the larger range and an event that occurred with probabilty η. This observation is then ignored when estimating future ranges. However if such extreme observations, extreme when evaluated with respect to the estimate, occur L times within the K periods then the statistician changes her estimate and includes these data points in the calculation of the estimated range of errors. Once again, this was a description of how the small error range is estimated. Whether or not the null hypothesis is rejected depends on the relationship between the realized data and these estimates.

We present our approach more formally. Beliefs about error ranges are determined as follows. For each t compute F_{t-1} measurable estimates a_t, b_t and c_t of a, b and c. We use these to estimate the errors $\hat{\varepsilon}_k$ that occured in the last K rounds, so $\hat{\varepsilon}_k = y_k - b_t y_{k-1} - (a_t + c_t k)$ for $k = t-K, ..., t-1$. We now compute the range of the estimated errors $[\bar{\varepsilon}_{t1}, \bar{\varepsilon}_{t2}]$ where $\bar{\varepsilon}_{t1} = \min_{k\in\{t-K,..,t-1\}} \{\hat{\varepsilon}_k\}$ and $\bar{\varepsilon}_{t2} = \max_{k\in\{t-K,..,t-1\}} \{\hat{\varepsilon}_k\}$. Finally we inflate this range by the factor φ to obtain the small range $[g_{1t}, g_{2t}] = [\varphi\bar{\varepsilon}_{t1}, \varphi\bar{\varepsilon}_{t2}]$ and large range $[\rho g_{1t}, \rho g_{2t}]$ of possible errors.

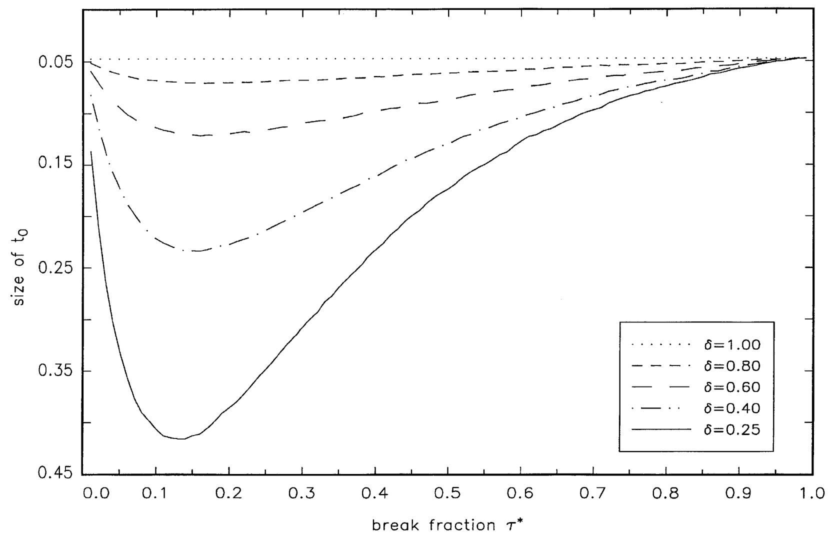
The constraints imposed by the large bounds are tested as follows. If for each choice of a_0 and c_0 there is some value such $y_t - b_0 y_{t-1} - a_0 - c_0 t \notin [\rho g_{1t}, \rho g_{2t}]$ then reject the null hypothesis. The test for the small bounds proceeds as follows. For each value of a_0 and c_0 consider how many values of $t \in \{1, ..., T\}$ there are such that $y_t - b_0 y_{t-1} - a_0 - c_0 t \notin [g_{1t}, g_{2t}]$. If for each choice of a_0 and c_0 there are too many values with this property, where too many is determined by the binomial test with success probability η and size α_B, then reject the null hypothesis.

The test depends on several tuning parameters. In particular, we calculate the test statistic for the sample excluding the first $\underline{T}$ observations. Out of the remaining observations, we impose $\bar{b}$ for the next $T_0 - \underline{T}$ observations and finally estimate for model (a) (model (ac)) a and b (a, b and c) for the remaining $T - T_0$ observations.

6 Simulations

6.1 Dickey-Fuller

The most popular unit root test, the Dickey Fuller test (DF, (Dickey and Fuller, 1979)), test is known to be oversized in several cases. First, the DF test crucially depends on a constant variance. Kim et al. (2002) have shown that the size can amount to 40% if there is a variance break early in the sample. They consider situation in which the first $\tau 100\%$ of observations have a constant variance σ^2 and the following $(1-\tau)\, 100\%$ observation have a variance of $\delta^2\sigma^2$. Figure 6.1 (taken from Kim et al, 2002) shows the size of the DF test as function of the break fraction τ^* and ratio of standard deviations before and after the break δ.



Size of DF when there is a break in the innovation variance

Second, the DF test is oversized when the errors depend on the lagged variable. Consider

$$\varepsilon_t = \upsilon_1(\upsilon_2 + \gamma y_{t-1})$$

where υ_1 is distributed as U[-$\frac{1}{2}$,$\frac{1}{2}$] and υ_2 is distributed as U[0,1]. Table 1 shows the size for several values of γ for $T = 200$.

Table 1: Size of DF when innovation depends on Y_{t-1}

γ	0	0.1	0.5	0.7	1
$a = 0$	5%	5.3%	7.5%	15.7%	41%.

Finally, Gospodinov and Tao (2011) show that the DF test can be highly oversized if the process has near integrated GARCH(1,1) errors as under the following specification

$$\begin{aligned} y_t &= b y_{t-1} + \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t} \eta_t \\ h_t &= \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \end{aligned}$$

with $\eta_t \sim iid\,(0,1)$ and $\alpha + \beta$ close to 1. The size exceeds 40% in the case that $\alpha + \beta$ is close to 1 as seen in Table 2.

Table 2: Size of DF for processes with GARCH(1,1) erorrs

$\alpha \setminus \beta$	0.3	0.399
0.6	9.3%	41.9%
0.65	12.4%	/
0.69	16.9%	/
0.699	43.6%	/

Given the above limitations of the DF, several extensions have been proposed (see e.g. Kim et al. (2002), Kim et al (2013), Gospodinov and Tao (2011)). However, only a few simulations are available, hence it is open whether these control the type I error.[7] Given a lack of formal understanding of these methods, we proceed with exact tests which due to their property of being exact control the type I error.

6.2 Wright

Wright (2000) proposes two exact versions of variance ratio tests: a rank test that assumes conditional homoskedasiticity and a sign test that allows for conditional heteroskedasticity. Because of the mentioned importance of conditional heteroskedasticity in economic and financial applications we proceed with the latter. The test assumes mean and median 0, therefore some weak form of symmetry. It tests whether the errors follow a martingale difference sequence, have median 0 and have a sign that is *iid*. Therefore, the test has power against different types of serial dependence coming from both autoregressive and long memory time series model. In case of a rejection, it is not obvious which of the three tested features is violated. Wright (2000) shows that his test has good size and power properties. The test is however not applicable when the data is asymmetric. In particular, Wright's sign test can be dramatically oversized when the innovations ε_t are asymmetric. To illustrate this, we

[7] A general feature of nonparametric models is that one cannot establish properties under the null hypothesis with simulations, as the set of data generating processes is too rich.

Table 3: Size of Wright's unit root test for processes with asymmetric errors

f	0.5	0.75	1	1.25	1.5	2
$median$	0.125	0.094	0	-0.1	-0.17	-0.25
$T = 200$	78.6	8.9	3.3	5.1	22.7	79.0
$T = 500$	99.3	15.7	3.9	7.9	42.9	98.6
$T = 1000$	100	27.4	3.4	12.0	71.8	100

use a simple example of an asymmetric data generating process, with density $\frac{f}{e(e+f)}$ between $[-e, 0]$ and density $\frac{e}{f(e+f)}$ between $[0, f]$. While the mean of this process remains zero, $Prob(\varepsilon_t > 0) = \frac{e}{f(e+f)}$ differs from 1/2 as long as $e \neq f$ and consequently the median equals $\frac{e(f-e)}{2f}$ if $f > e$ and $\frac{f(e-f)}{2e}$ if $f < e$. Table 3 displays, for $e = 1$ and several values for f, the median and the rejection probabilities for 1000 simulations and sample sizes $T = 200$, 500 and 1000. Thus Wright's sign test does not control the type I error when the median 0 assumption is not met. Since asymmetry can be a very relevant feature, especially of financial data, this constitutes a major drawback of the test.

6.3 Luger

Luger (2003) proposes a test that applies when errors are symmetrically distributed and have no point masses at 0. This test uses long differences $z_t = \Delta y_{t+m} - \Delta y_t$ in order to eliminate the drift term and applies a sign test. Under the null of a unit root, $z_t = \varepsilon_{t+m} - \varepsilon_t$ is symmetric with median 0, while under the alternative, the median of z_t is a function of a and b and differs from 0. Luger (2003) proposes further an approximate test which albeit it is not exact.

Luger's exact test shows very high power for stationary processes with autoregressive parameters very close to unity, even for relatively small sample sizes. However, apart from the low power when the level parameter a is close to 0, the test lacks power in the following situations: autoregressive parameters (slightly farer) away from unity, the series starts prior to the observed sample and the sample size increases as b is held constant.

First, we consider the sign statistic S_m for sample size $T = 200$ and for several values of b, replicating table 2 in Luger (2003), however now including values of b below 0.96. As can be seen in Table 4, for values of b relatively close to one, the power is decreasing rather than increasing. The last two rows illustrate that for uniformly distributed errors U[-1,1] the power is higher but suffers still under the same problem.

Table 4: Rejection probability of Luger's unit root test for processes farer from unit root

	$T\backslash\ b$	1	0.99	0.98	0.97	0.96	0.94	0.92	0.9	0.8	0.7
normal	200	3.7	66.8	64.6	40.7	25.6	11.2	5.1	2.8	0.4	0.7
	500	3.9	93.2	45.1	16.7	9.0	2.7	1.2	0.8	0.3	0.3
$uniform[-1,1]$	200	4.9	97.4	97.6	87.4	64.7	30.1	13.1	6.7	0.7	0.3
	500	2.8	100	92.2	54.7	25.6	7.3	3.3	2.1	0.7	0.3

This constitutes a clear weakness of the test, since then many (persistent) stationary series cannot be distinguished from a unit root. E.g. for $T = 500$ and normally distributed error, only series with $0.96 < b < 1$ can be distinguished from unit root processes.

Second, Luger´s high power crucially depends on the fact that for stationary processes with b close to unity starting at $y_0 = 0$, the process takes some time to reach the mean $\frac{a}{1-b}$, which is large for the cases considered in Luger. Once it has reached this mean due to its stationarity it moves around it. If we let the observed series start close to the unconditional mean, Luger´s power drops considerably. Table 5 shows the power for $T = 200$ and the series generated with normally distributed errors with $b = 0.97, 0.98$ and 0.99 and and T_{ps} observations prior to the observed sample.

Table 5: Power of Luger's exact test for processes with T_{ps} presample observations

T_{init}	0	10	50	100	200	500
$b = 0.99$	66.8	58.5	31.4	14.3	5.4	3.3
$b = 0.98$	64.6	43.2	11.5	4.3	1.4	1.5
$b = 0.97$	40.7	24.0	4.1	1.1	1.2	1.8

If we let the series start at its unconditional mean $y_0 = \frac{a}{1-b}$ the power will be close to the one in the last column. This is a big drawback, since there is no reason why an observed stationary series would start at a value far from its unconditional mean. Note that this is not a problem of an initial condition as discussed e.g. in Müller and Elliot (2003) in the unit root literature, since here the first observation is bounded in probability rather than of order $O(T^{1/2})$.

Further, since Luger obtains his power from using some local to unity behavior, the tests are not consistent in the sense that with increasing sample size the power converges to 1 as can be seen in Table 6.

This is because as the sample size increases it is as if b was farer away from a unit root and as previously mentioned the power would decrease.

Luger's test is designed for symmetric errors. Note that the errors that need to be symmetric are $\eta_t = \varepsilon_{t+T/2} - \varepsilon_t$ rather than ε_t. Therefore, Luger still works if ε_t is

Table 6: Power of Luger's exact test for processes with larger sample sizes

$b\backslash\ T$	100	200	500	1000	2000	5000
1	1.9	3.6	4.1	4.6	5.0	4.3
0.99	21.8	66.5	92.8	83.9	95.0	54.2
0.98	33.6	63.0	46.2	21.4	29.0	8.7
0.97	30.2	39.4	18.1	9.7	9.0	3.2

Table 7: Size of Luger's unit root test for processes with asymmetric errors

f	1	1.25	1.5	2	2.5	3
$T = 200$	3.5	4.1	7.2	16.4	29.5	44.3
$T = 500$	3.8	5.2	10.5	32.7	60.9	80.6

drawn from a time constant asymmetric distribution, a feature that interestingly is not noted in Luger. However, for more general asymmetric errors it fails to be exact. In the following, we analyze how crucial the symmetry assumption is by looking at an error distribution with the asymmetry changing over time. We generate data from a mean zero distribution which has density $\frac{f}{e(e+f)}$ between $[-e, 0]$ and density $\frac{e}{f(e+f)}$ between $[0, f]$ and where f equals e in the first half of the sample and equals γe in the second half. It is easily computed that $P(\eta_t > 0) \neq 0$. Table 7 displays the type I error probabilities for $T = 200$ and $NS = 10,000$ and different values of γ. This table illustrates that without symmetry the Luger exact test can be dramatically oversized.

6.4 Our test

We now present simulations for our test. First, we consider the performance of the proposed test as a unit root test. For this we consider two different cases:

(1) $\mathrm{H}_0 : b_0 = 1, a = 0$ vs $\mathrm{H}_1 : b < 1, a \neq 0$

(2) $\mathrm{H}_0 : b_0 = 1, a \neq 0, c = 0$ vs $\mathrm{H}_1 : b < 1, a \neq 0, c \neq 0$

In (1), the process has a unit root but no linear trend under the null and is stationary under the alternative. In (2), the process has a unit root plus a linear trend under the null and is trend stationary under the alternative. For (1), we generate the data as

$$y_t = a + by_{t-1} + \varepsilon_t,$$

with $b \leq 1$ and $\varepsilon_t \sim U[-1, 1]$, implying a true $g = 1$, and test $\mathrm{H}_0 : b_0 = 1, a = 0$. First, note that if the data is generated as in Luger with $y_0 = 0$, we find high power very close to $b = 1$ for smaller values of g as illustrated in Table 8 for $T = 200$ and different values of b and g.

Table 8: Power of our test when $y_0 = 0$

$g\backslash b$	0.9	0.925	0.95	0.975	1
1	100	100	100	100	0
1.05	98	100	99	100	0
1.15	75	72	95	97	0
1.25	17	26	44	63	0
1.35	1	3	5	9	0

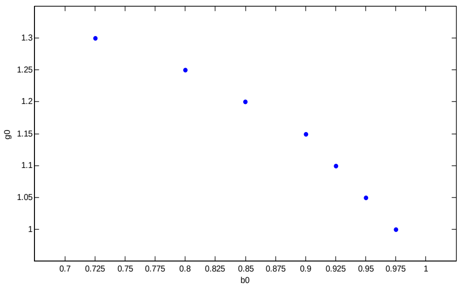
However, as discussed before, we consider this as a rather unnatural scenario and, thus, we consider in the following processes that start close to their unconditional mean. Therefore, for $b < 1$, we choose a so that $a/(1-b) = 1$ in order to mantain the same mean over different values of b. For $b = 1$, we choose $a = 0$.

We calculate the rejection probabilities of the null for values of several values of g and for $T = 200$, 500 and 1000. [8]

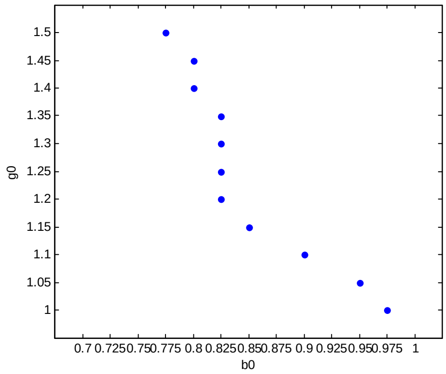
Figure 1 shows for 1000 simulations for each true b, the assumed bound g_0 at which in more than 50% of the cases the null $b_0 = 1$ is rejected with the parameters $\underline{\mathrm{T}}= 10, T_0 = 200$, $\bar{b} = 0.8$ for $T = 200$ and 500 and $\bar{b} = 0.85$ for $T = 1000$.

[8] The simulations are based on a previous version, in which we reject if more than 1% of the observations lie outside of a band with width $2g$. Since the errors here are uniformly distributed and therefore extremer observations occur relatively likely the results should be similar. The power for a given combination of g_0 and b_0 would be lower because first more outliers would be allowed under the null and rather than dropping the outliers, we would include them together with a larger range entering in the test.

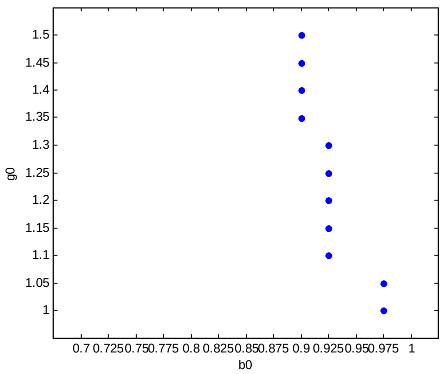
Figure 1: Model (a). Unit root test. 50 percent rejection probabilities as function of b and g0



i) $T = 200$



ii) $T = 500$



iii) $T = 1000$

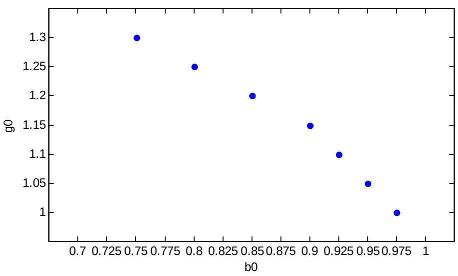
For larger sample sizes, the power of the procedure is relatively high. For $T = 500$ for example, we reject in over 50% of the cases for $b = 0.825$ if we assume an error bound g of up to 1.35. For $T = 1000$, we reach the 50% rejection probabilitity already for $b = 0.925$ for similar error bound assumption.

For (2), we generate the data as

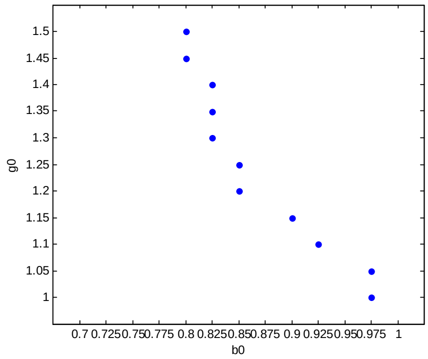
$$y_t = a + ct + by_{t-1} + \varepsilon_t$$

with $b < 1, a = 1$ and $\varepsilon_t \sim U[-1, 1]$ and test $\mathrm{H}_0 : b = b_0 = 1,\ c = c_0 = 0$. For $b < 1$, we choose c so that $c/(1 - b) = 1$ in order to maintain the slope constant over b. For $b = 1$, we choose $c = 0$. The test again depends on several tuning parameters. As in the previous case we choose $\underline{\mathrm{T}} = 10, T_0 = 200, \bar{b} = 0.8$ for $T = 200$ and 500 and 0.85 for $T = 1000$. Figure 2 shows for each g_0 the true b at which we reject in at least 50% of the cases for 1000 simulations.

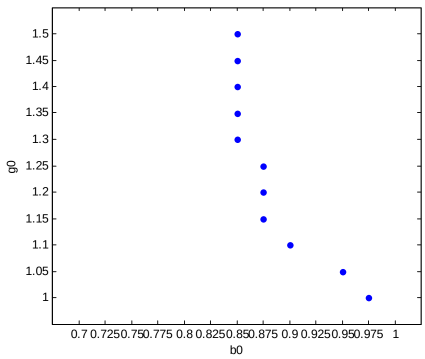
Figure 2: Model (ac). Unit root test. 50 percent rejection probabilities as function of b and g_0



i) $T = 200$



ii) $T = 500$



iii) $T = 1000$.

Clearly, the smaller the assumed g_0, the smaller is the confidence interval for b. The confidence intervals shrink for each g_0 in the sample size. The width of the confidence intervals for the case without and with trend are similar. It is clear that the results crucially depend on the assumed g_0. To obtain a specific confidence interval, one can use the adaptive range approach.

6.4.1 Other error distributions: Truncated Normal, Asymmetric errors and changing variance

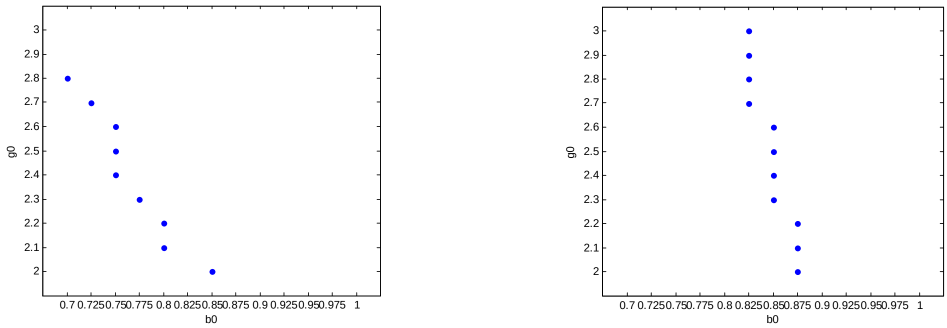
Here, we consider alternative error distributions for the test without a time trend. The one with a time trend behaves similarly. We consider the same tuning parameters as in the previous section and consider the case $T = 500$ and $T = 1000$ and 100 simulations. We consider

(i) ε_t follows a truncated (standard) normal with support $[-2, 2]$

(ii) ε_t follows the asymmetric distribution considered in Section 6.2 with $e = 1$, $f = 1.5$, namely with density $\frac{3}{5}$ between $[-1, 0]$ and density $\frac{4}{15}$ between $[0, 1.5]$.

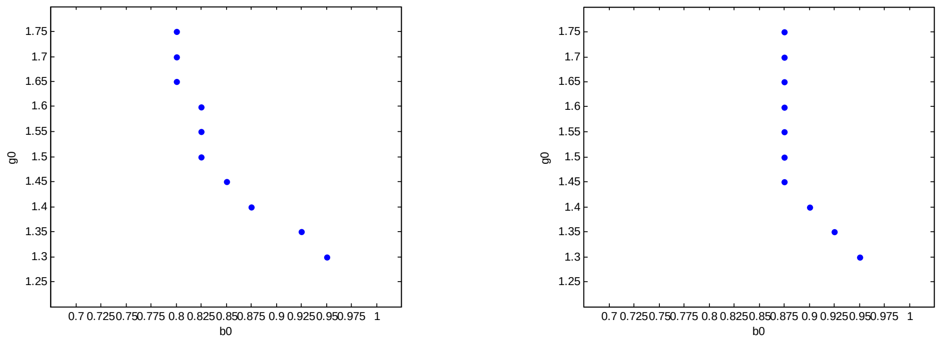
(iii) ε_t follows a uniform distribution in support $[-1, 1]$ in the first half and with support $[-0.5, 0.5]$ in the second half

Figure 3: Model (a). Test for different error distributions. 50 percent rejection probabilities as function of $b\S and\S g$

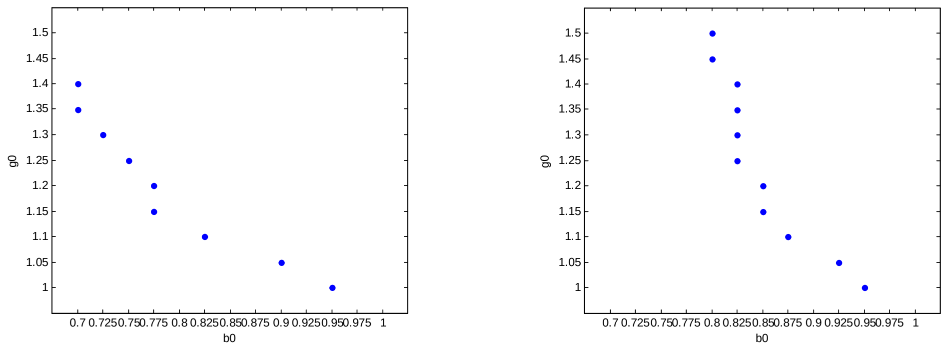


...

i) truncated normal: $T = 500$ $T = 1000$



ii) asymmetric distribution: $T = 500$



iii) changing variance. $T = 500$ $T = 1000$

First, (i) illustrates that in cases in which the bounds are away from the region where the larger mass of the errors lies, the power decreases. In this case, similar to the uniform examples, the bounds are too large when compared to the variation of the variable.

Our procedure applies straightforwardly to the asymmetric example in (ii) since only the width of the support of ε_t is used in constructing the bounds in the test.

The changing support/variance example decreases the power since now constant bounds are too wide for the second part of the sample. Therefore, the power decreases in the ratio of the variances. Notice that if only the variance changes but not the bounds, the power would be larger and would lie somewhere between the uniform cases and case (i) in this section.

Finally, our procedure with constant bounds does not behave well when the bounds change drastically as in the case of GARCH errrors

(iv) $\varepsilon_t = exp\,(h_t/2)\,\eta_t$ where $h_t = 0.95h_{t-1} + \zeta_t$ with ζ_t iid $N\,(0, 1/10)$ and η_t $U[-1, 1]$.

The reason is that a bound large enough to cover all errors can be clearly too big for the largest part of observations. Currently, we are working on applying the adaptive approach to this type of data.

6.4.2 Confidence intervals for b

In this section, we simulate the confidence intervals when testing the null of a stationary AR(1) versus the alternative of a unit root process. We again distinguish the cases (1) without trend and (2) with trend. For (1), we generate data as

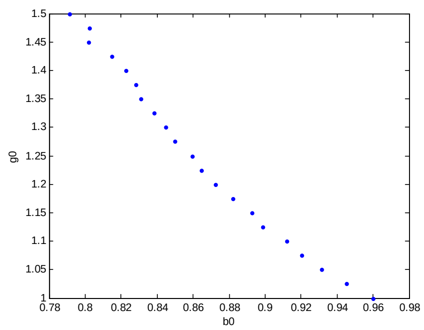
$$y_t = y_{t-1} + \varepsilon_t$$

and test $H_0 : b = b_0$ with $b_0 \in [0, 1]$ in the model

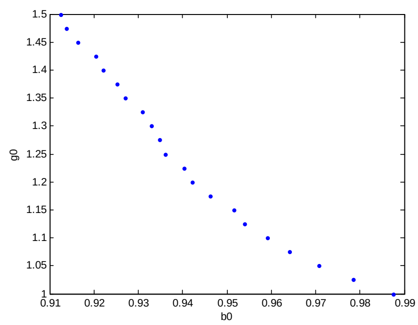
$$y_t = a + ct + by_{t-1} + \varepsilon_t.$$

We obtain a confidence set for b and g for which we cannot reject the null of H_0 : $b = b_0$ and $g = g_0$. Figure 4 displays for each g the average (over 1000 simulations) confidence interval lower bounds, i.e. the confidence interval for a given g is the region on the right of the dots.

Figure 4: Confidence set for b_0 and g_0 for processes without linear trend



i) $T = 200$



ii) $T = 500$

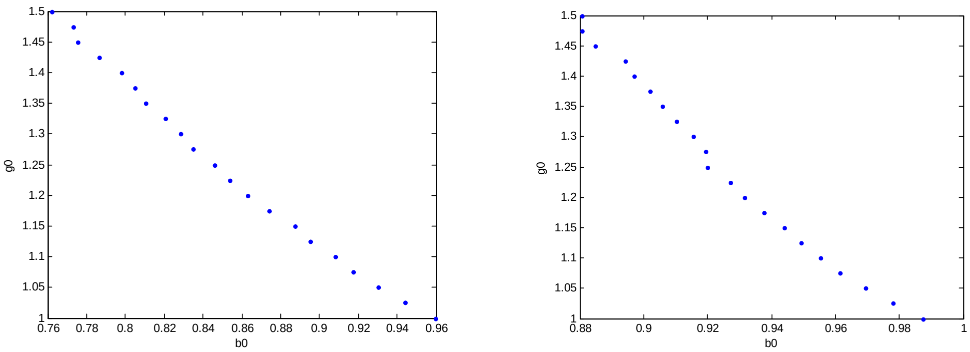
For (2), we generate data as

$$y_t = a + y_{t-1} + \varepsilon_t$$

and test several $H_0 : b = b_0$ in the model

$$y_t = a + ct + by_{t-1} + \varepsilon_t.$$

Figure 5: Confidence set for b_0 and g_0 for processes with linear trend



i) $T = 200$ ii) $T = 500$

The confidence sets are quite narrow, implying that for not too high values of g_0, the confidence intervals are narrow. This results from the fact that nonstationary process with given length T leaves any bound with high probabilitly and the weighted averages that are compared in the relevant test often differ by bigger amounts. The confidence intervals for given g are slightly smaller in model (2). They shrink with the sample size T.

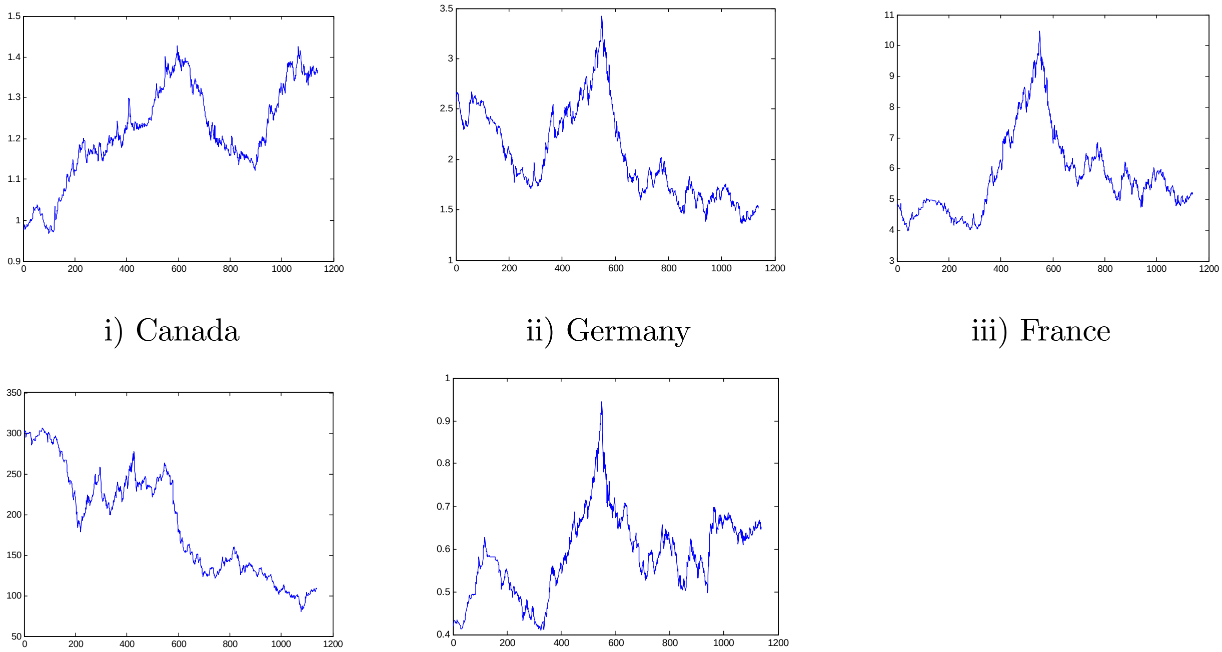
In this section we have seen that the proposed procedure has reasonable power for not too small sample sizes, when the error bounds do not change too much and when the tails are not too thin. Especially, the confidence intervals for processes with unit roots can be narrow potentially providing strong evidence in favor of highly persistent series.

7 Empirical application

7.1 Wright example of exchange rates

Next, we apply our proposed testing procedure to the time series of exchange rate returns analyzed by Wright (2000). As in Wright (2000), we consider weekly nominal exchange rates for the Canadian dollar, French franc, German mark, Japanese yen, and British pounds, all relative to the US dollar. The data ranges from August 7,1974 to May 29, 1996 and is constructed as the differences of the log exchange rate. Figure 6 displays the five exchange rate series.

Figure 6: Exchange rates for Canada, Germany, France, Japan, and UK relative to US



i) Canada ii) Germany iii) France

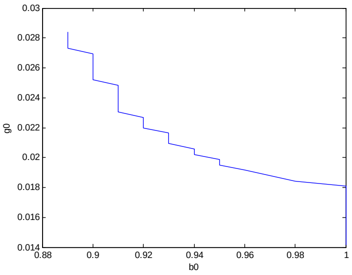
iv) Japan v) UK

First, we test for a unit root. Since under the alternative of a stationary series, there would be a linear trend, we work with model (ac). As discussed, we have to choose several tuning parameters. For several sets of tuning parameters and for all values of g, we cannot reject the null of a unit root for any of the five series. The reason is that the estimates of b are very close to 1 in all cases.

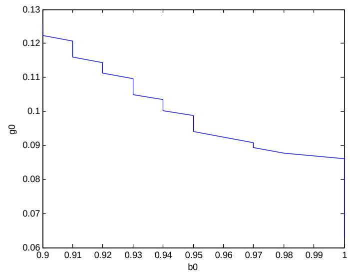
Next we construct the confidence sets for b and g for all five series within the model (ac). We start the testing procedure from $\underline{T}$= 100, we choose the probability for the ocurrence of an outlier to be $\eta = 1\%$ and the ratio of small to big range $\rho = 3$. We further divide the significance level of the two tests, inequality and outlier, into $\alpha_1 = 2.5\%$ and $\alpha_2 = 2.5\%$.

Figure 7 displays these sets for values of several values of g. It can be seen that for $b_0 = 1$, the smallest bounds are admissable.

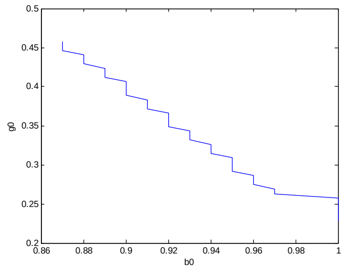
Figure 7: Confidence sets for the autoregressive parameter and the bound g_0 for the exchange rates within model (ac)



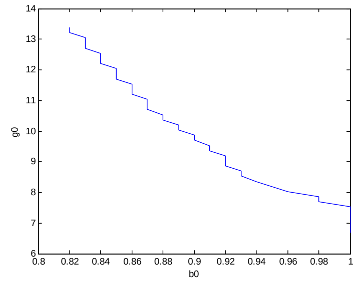
i) Canada



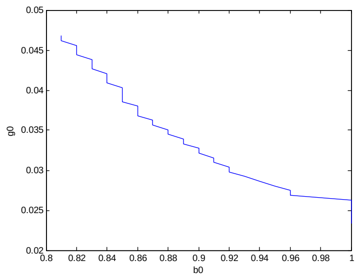
ii) Germany



iii) France



iv) Japan



v) UK

Next, we apply the approach with the adaptive range and interval in which the observations should lie under the null. For this test, we additionally choose the window size in which we estimate the bounds $K = 100$ and the possible increase of the small range $\varphi = 1.2$. Table 9 displays the confidence sets for the estimates of the autoregressive parameter for the five countries.

Table 9: Confidence intervals for the autoregressive parameter for the exchange rates for the adaptive approach within model (ac)

	Canada	Germany	France	Japan	UK
CI	$[0.96, 1]$	$[0.94, 1]$	$[0.97, 1]$	$[0.93, 1]$	$[0.94, 1]$

First, the fact that $b_0 = 1$ lies in the confidence intervals confirms that the unit root cannot be rejected for any of the series. Second, the confidence sets are relatively narrow providing evidence for a high persistence of the series. Clearly for larger values of φ the confidence intervals increase. They are not overly sensitive to the choice of α_1 and α_2.

Our results are somehow in contrast to the results in Wright (2000) who rejects the martingale hypothesis. However, notice that it is also possible that his rejection comes from the fact that the true median differs from zero.

8 Conclusion

We consider tests for the autoregressive coefficient in a simple autoregressive model in which errors have mean 0. We present the first explicit test that allows for heteroskedasticity and asymmetric errors (the conditional median need not be equal to 0). The possibility to construct such a test by evaluating a martingale comes at no surprise given the inequality of Hoeffding (1963) and the "trick" of putting the lagged variable on the left hand side (as used for instance in (Dickey and Fuller, 1979)). Given the nonexistence result of Dufour (2003), one has to impose some type of bounds on the summands in the martingale. The power of such a test will depend on how tight one can choose these bounds. Here the inequality of van de Geer (2002) plays a key role. It allows for a flexible approach as bounds may depend on past information, in contrast to the inequality of Hoeffding (1963) where the bounds need to be determined before gathering the data. We substantially improve the inequality of van de Geer (2002) by including the aggregate measure of the ranges, C_T^2, in the test statistic as opposed to assuming some exogenous bound on this expression. Equipped with this tool we proceed, motivated by the Dickey-Fuller test, by writing up a sum of terms where each term is the product of the past and the future and has mean 0. By optimizing the term that involves the past we obtain an expression that describes how precise one can predict the future. The cutoff in the probability bound reflects how difficult it is to predict the future under the null hypothesis where it is uncorrelated with the past. However, as one moves away from the null, past and future become correlated, and the power of the test increases as the future becomes easier to predict. Difficulties however arise when one gets too close to the case of a unit root where the future is hard to predict with high accuracy. To accommodate for alternatives that have a unit root we augment the above with an inequality that is based on taking differences, relying on the nonstationarity of the sequence of differences.

For simplicity we only present a model with a single lag, the method directly extends to multiple lags. One moves all lagged variables to the left hand side, includes them in the definition of W and then proceeds as above. This yields a simultaneous test of all lagged variables (the value of each lagged variable to be tested is included in the null hypothesis) and translates into a confidence set for the set of lagged variables. Confidence intervals for any specific lag are derived via projection, considering all values that are contained in some vector that belongs to the confidence set.

Our test of the autoregressive coefficient b can be easily extended to allow for exogenous regressors. For stationary alternatives, adjust the coefficients τ to create an unbiased estimate of W_t by using the future Ws. For nonstationary alternatives replace the differences with some weighted sum of the Ws that has mean 0. Tests of coefficients a and c within the times series model are also easy to construct. Select a test for the case where b is known and then, following Campell and Dufour (1997), run this test with level α_2 for all values of b that belong to the $1-\alpha_1$ confidence interval of b, where $\alpha_1+\alpha_2=\alpha$. Which test to select when b is known and errors are only constrained to follow a martingale difference sequence will be left for future research. For independent errors one can directly apply the methods in Gossner and Schlag (2013). This paves the road for a nonparametric analysis of dynamic panels.

More work is needed to understand how best to deal with the bounds. Following Dufour (2003), one cannot test the martingale hypothesis with a non trivial test. Additional constraints on the data generating process are needed. In this paper we show how one can test the martingale hypothesis for given bounds on the errors. An alternative followed among others by Wright (2000) and Luger (2003) is to test the intersection of the martingale hypothesis with the hypothesis that all median errors are equal to 0.[9] The need of constraints is a general phenomenon of mean testing, first identified by Bahadur and Savage (1956) for testing the mean of an i.i.d. sequence. We add a bound on the errors which is much less restrictive than imposing that the conditional median errors are equal to 0. This can be seen indirectly by observing that it is not conceivable to derive the distribution of a meaningful test statistic in our model while this is done in Wright (2000) and Luger (2003). This can also be seen directly. We explain intuitively, the topological details are easily added but do not add any insight. Consider only models in which errors follow a martingale difference sequence. Consider any model that additionally has the median property, namely where all errors have median 0. Then almost all small changes of the model destroy the median property. The set of models with the median property is very fragile within the set of models. The opposite is true for almost all models that satisfiy the bounded error property as modelled in this paper. For almost any model with the bounded error property, all sufficiently close models also have this property.

The cleanest way to embedd the bounds in the results is to consider the 95% confidence set that includes both bounds and autoregressive coefficients. To the user who wishes to have a unique recommendation, reject or not, we suggest an approach of how to specify how future ranges depend on past observations.

There are obvious alternatives for dealing with the bounds. One may wish to choose the bounds from earlier studies on similar data sets. One could make some assumptions on the tails of the errors to be able to estimate the bounds based on the data. In view of heteroskedasticity we allowed the bounds to change over time.

In general, the fact that bounds are never observable makes them an intriguing topic for future research.

---

[9] As errors are heteroskedastic this adds a number of equality constraints that is equal to the size of the data set.

References

[1] Bahadur, R. R., and L. J. Savage (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *The Annals of Mathematical Statistics*, 27, 1115–1122.

[2] Belaire-Franch, J., and D. Contreras (2004): "Ranks and Signs-Based Multiple Variance Ratio Tests," Working paper, University of Valencia.

[3] Campbell, B., and J.-M. Dufour (1997): "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter," *International Economic Review*, 38, 151–173.

[4] Charles, A., and O. Darné (2009): "Variance-Ratio Tests of Random Walk: An Overview," *Journal of Economic Surveys*, 23(3), 503–527.

[5] Dickey, D. A., and W. A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

[6] Dufour, J.-M. (2003): "Identification, Weak Instruments, and Statistical Inference in Econometrics," *The Canadian Journal of Economics/Revue Canadienne d'Economique,* 36, 767–808.

[7] Gospodinov, N., and Y. Tao (2011): "Bootstrap Unit Root Tests in Models with GARCH(1,1) Errors," *Econometric Reviews*, 30(4), 379–405.

[8] Gossner, O., and K.H. Schlag (2013): "Finite-Sample Exact Tests for Linear Regressions with Bounded Dependent Variables", *Journal of Econometrics,* 177, 75–84.

[9] Hoeffding, W. (1963): "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58, 13–30.

[10] Kim, B.H., T.-H. Kim, H.-H. Moon, and S.-B. Jeong (2012): "Unit Root Test in the Presence of Multiple Breaks in Variance," http://www.fas.nus.edu.sg/ecs/scape/doc/25Jan13/A2-2.pdf.

[11] Kim, T.-H., S. Leybourne, and P. Newbold (2002): "Unit Root Tests with a Break in Innovation Variance," *Journal of Econometrics*, 109(2), 365–387.

[12] Kim, J. H., and A. Shamsuddin (2008): "Are Asian stock markets efficient? Evidence from new multiple variance ratio tests," *Journal of Empirical Finance*, 15, 518–532.

[13] Luger, R. (2003): "Exact Non-Parametric Tests for a Random Walk with Unknown Drift under Conditional Heteroscedasticity," *Journal of Econometrics*, 115, 259–276.

[14] Müller, U. K, and G. Elliott (2003): "Tests for Unit Roots and the Initial Condition," *Econometrica*, 71(4), 1269–1286.

[15] Pinelis, I. (2006): "On Normal Domination of (Super)Martingales," *Electronic Journal of Probability*, 11, 1049–1070.

[16] van de Geer, S. (2002): "On Hoeffding's Inequality for Dependent Random Variables," in: *Empirical Process Techniques for Dependent Data*, ed. by H. Dehling, T. Mikosch and M. Sorensen, pp. 161–170, Birkhauser, Boston.

[17] Yates, F. (1934): "Contingency Tables Involving Small Numbers and the χ^2 Test," *Supplement to the Journal of the Royal Statistical Society* 1, 217–235.

9 Appendix

Consider a random variable X with the following properties. There exist $L_1, L_2, U_1, U_2 \in \mathbb{R}$ with $L_1 < U_1$ and $[L_1, U_1] \subset [L_2, U_2]$ and $\eta \in (0,1)$ such that $EX = 0$, $P(X \in [L_2, U_2]) = 1$ and $P(X \in [L_1, U_1]) = 1 - \eta$. We wish to bound $E(\exp(\beta X))$.

Since e^x is convex the worst case is attained when all mass is shifted to the boundary, hence it is enough to consider X that has support in $\{L_1, L_2, U_1, U_2\}$. Let $\lambda = P(X = L_1)$ and $\mu = P(X = L_2)$. Then $\lambda \in [0, 1-\eta]$ and $P(X = U_1) = 1-\eta-\lambda$, similarly $\mu \in [0, \eta]$ and $P(X = U_2) = \eta - \mu$. Moroever, $EX = \lambda L_1 + (1-\eta-\lambda) U_1 + \mu L_2 + (\eta - \mu) U_2 = 0$, hence

$$\lambda = \bar{\lambda}(\mu) := \frac{U_1 - (U_2 - L_2)\mu + \eta(U_2 - U_1)}{U_1 - L_1}.$$

Consequently,

$$E\left(e^{\beta X}\right) \leq \bar{\lambda}(\mu) e^{\beta L_1} + \left(1 - \eta - \bar{\lambda}(\mu)\right) e^{\beta U_1} + \mu e^{\beta L_2} + (\eta - \mu) e^{\beta U_2}.$$

In particular the right hand side is linear in μ which makes it easy to maximize.

In fact, under simple constraints on η we find that the maximum is attained when $\mu \in \{0, \eta\}$. Namely, we find that

$$\bar{\lambda}(\mu) \geq \frac{U_1 - (U_2 - L_2)\eta + \eta(U_2 - U_1)}{U_1 - L_1} = \frac{(1-\eta) U_1 + \eta L_2}{U_1 - L_1} \geq 0$$

if $\eta \leq \frac{U_1}{U_1 - L_2}$ and

$$\bar{\lambda}(\mu) \leq \frac{U_1 + \eta(U_2 - U_1)}{U_1 - L_1} \leq 1 - \eta$$

if $\eta \leq \frac{-L_1}{U_2 - L_1}$.

Proposition 7 *If* $\eta \leq \min\left\{\frac{U_1}{U_1-L_2}, \frac{-L_1}{U_2-L_1}\right\}$ *then*

$$E\left(\exp\left(\beta X\right)\right) \leq \max\left\{\begin{array}{c} \frac{U_1+\eta(U_2-U_1)}{U_1-L_1}e^{\beta L_1} + \frac{-(1-\eta)L_1-\eta U_2}{U_1-L_1}e^{\beta U_1} + \eta e^{\beta U_2}, \\ \frac{U_1-(U_2-L_2)\eta+\eta(U_2-U_1)}{U_1-L_1}e^{\beta L_1} + \frac{-(1-\eta)L_1+(U_2-L_2)\eta-\eta U_2}{U_1-L_1}e^{\beta U_1} + \eta e^{\beta L_2} \end{array}\right\}.$$