

Reese, Simon; Li, Yushu

Working Paper

Testing for Structural Breaks in the Presence of Data Perturbations: Impacts and Wavelet Based Improvements

Working Paper, No. 2013:36

Provided in Cooperation with:

Department of Economics, School of Economics and Management, Lund University

Suggested Citation: Reese, Simon; Li, Yushu (2013) : Testing for Structural Breaks in the Presence of Data Perturbations: Impacts and Wavelet Based Improvements, Working Paper, No. 2013:36, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/260092>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Working Paper 2013:36

Department of Economics
School of Economics and Management

Testing for Structural Breaks in the Presence of Data Perturbations: Impacts and Wavelet Based Improvements

Simon Reese
Yushu Li

October 2013



LUND
UNIVERSITY

Testing for structural breaks in the presence of data perturbations

---- Impacts and wavelet based improvements

Simon Reese¹ and Yushu Li²

Abstract

This paper investigates how classical measurement error and additive outliers influence tests for structural change based on F -statistics. We derive theoretically the impact of general additive disturbances in the regressors on the asymptotic distribution of these tests for structural change. The small sample properties in the case of classical measurement error and additive outliers are investigated via Monte Carlo simulations, revealing that sizes are biased upwards and that powers are reduced. Two wavelet based denoising methods are used to reduce these distortions. We show that these two methods can significantly improve the performance of structural break tests.

JEL classification: C11, C12, C15

Keywords: Structural breaks, measurement error, additive outlier, wavelet transform, empirical Bayes thresholding

1. Introduction

The detection of structural breaks is an integral part of economic time series analysis. If an econometric model is to reflect the optimal decision rules of individuals, its structure will be altered by changes in the environment of the economic agents that are studied (Lucas, 1976). Ignoring the presence of a structural break in the sample will in general lead to a misspecification of the econometric model that renders the obtained coefficients useless for

¹ Department of Economics, Lund University, Sweden. E-mail: simon.reese@nek.lu.se

² Department of Business and Management Science, Norwegian School of Economics, Norway. Yushu Li gratefully acknowledges funding from Swedish Research Council (project number 421-2009-2663). The authors would like to thank David Edgerton for helpful comments.

testing or further statistical inference (Stock, 1994). It is well known that in practice economic data cannot be produced without some degree of imperfection or incompleteness (Morgenstern, 1963, p.39). Evidence of the presence of inaccuracies in economic data has been found repeatedly, see van Bergeijk (1995, p. 3). Balke and Fornby (1994) argue that rare influential observations are also present in most macroeconomic indicators and account for a significant share of the overall variance in the sampling period. The implications of both classical measurement error and additive outliers in the explanatory variables of an econometric model are well documented. As pointed out in *e.g.* Bound *et al.* (2001, p. 3712), classical measurement error leads to inconsistency of the OLS estimator. In the case of outliers, the estimated coefficients will be biased in finite samples (see *e.g.* Davidson and MacKinnon, 2009, Ch. 2.6).

As far as we know the only attempt to analyse how data imprecisions affect structural break tests is Rodrigues and Rubia (2011), who focus on CUSUM-type tests for detecting structural change in the variance. This paper extends these authors' focus on classical measurement error and additive outliers to the widely used Quandt-Andrews family of tests for structural breaks (Quandt, 1960; Andrews, 1993; Andrews and Ploberger, 1994). We first give a theoretical derivation that shows that the asymptotic distributions of the structural test statistics are in general influenced by additive stochastic perturbations in the regressors. Monte Carlo simulations for finite sample sizes then show that both the size and power of these tests will be distorted in the presence of either classical measurement error or additive outliers. Our results show that classical measurement error may reduce test power significantly while having a relatively small impact on the size. The presence of additive outliers greatly inflates sizes while only exerting a moderately negative effect on power. The empirical Bayesian thresholding procedure of Johnstone and Silverman (2004, 2005a,b) is then used to remove inaccuracies from the data. This wavelet based method is chosen due to its ability to obtain resolution in both the frequency and time domains, which makes it possible to separate characteristics of a given series into different scales that can be treated separately. A simulation study using the corrected data shows that the size and power distortions due to additive outliers can be greatly reduced and that the negative effects on test power arising from measurement errors with can largely be mitigated.

The paper is organized as follows. Section 2 investigates how measurement errors and additive outliers influence F -statistic based tests for structural breaks. Section 3 covers the removal of data inaccuracies via empirical Bayesian thresholding and the performance of the breakpoint test on the corrected data. Finally, section 4 concludes.

2. The impact measurement inaccuracies on F -statistic based tests for structural break

Apart from the testing procedure that employs the generalized fluctuation test framework such as the CUSUM tests (Brown *et al.* 1975; Kuan and Hornik, 1995), the literature on tests for structural change in linear regressions is dominated by F -test based approaches that can be lead back to Chow's "coefficients equality" test (Chow, 1960; Quandt, 1960; Hansen, 1992b; Andrews, 1993; Andrews and Ploberger, 1994). Consider the linear model $y_t = \beta_t' X_t + e_t$, $t = 1, \dots, T$ with y_t denoting the dependent variable and $X_t = (x_{1t}, \dots, x_{kt})'$ denoting the $k \times 1$ vector of stochastic or nonstochastic regressors. We test the null hypothesis of no structural break in the parameter vector $\beta_t = (\beta_{1t}, \dots, \beta_{kt})'$, against the alternative hypothesis H_1 of a discrete break point in period at time point m .

$$\begin{aligned} H_0: & \beta_t = \beta; \forall t = 1, \dots, T \\ H_1: & \begin{cases} \beta_t = \beta, & t \leq m \\ \beta_t = \beta + \gamma, & t > m \text{ and } \gamma \neq 0 \end{cases} \end{aligned}$$

Chow (1960) proposed the test statistic

$$F_T\left(\frac{m}{T}\right) = \frac{SSR_{1,T} - (SSR_{1,m} + SSR_{m+1,T})}{(SSR_{1,m} + SSR_{m+1,T}) / (T - 2k)}$$

when the potential break time m is known, where $SSR_{1,T}$, $SSR_{1,m}$ and $SSR_{m+1,T}$ denote the sums of squared residuals of regressions for the whole data set and the subsamples before and after the break date. The Quandt-Andrews family of tests (Quandt, 1960; Andrews, 1993; Andrews and Ploberger, 1994) relaxes the assumptions about prior knowledge of exact time m to an interval $[m_1, m_2]$ that contains the breakpoint. This family covers three test statistics, whose Wald test form is defined as:

$$\begin{aligned} SupF_T &= \max_{m \in [m_1, m_2]} F_T\left(\frac{m}{T}\right) \\ AveF_T &= \frac{1}{m_2 - m_1 + 1} \sum_{m=m_1}^{m_2} F_T\left(\frac{m}{T}\right) \\ ExpF_T &= \ln\left(\frac{1}{m_2 - m_1 + 1} \sum_{m=m_1}^{m_2} \exp\left(\frac{1}{2} F_T\left(\frac{m}{T}\right)\right)\right) \end{aligned}$$

To compare the performance of these three tests, Kim and Perron (2009) presented a comprehensive power investigation using finite samples. Their investigation assumes the exogeneity of the independent variable, *i.e.* $E[e_t | X_t] = 0$. In the case of additive

perturbations in the data, however, this assumption will not necessarily be satisfied. Let $y_t^* = y_t + v_t$ and $X_t^* = X_t + U_t$ denote the observed values of the given data, with additive perturbations v_t and the $k \times 1$ vector U_t . This definition of v_t and U_t nests both classical measurement error and additive outliers, *i.e.* rarely occurring extreme values that cannot be explained by the DGP of the respective variables. However, distinguishing measurement errors from outliers entails further assumptions to be made about the stochastic properties of v_t and U_t . Substituting $y_t = y_t^* - v_t$ and $X_t = X_t^* - U_t$ into the true linear relation $y_t = \beta_t' X_t + e_t$ yields $y_t^* = \beta_t' X_t^* + e_t^*$, where $e_t^* = e_t + v_t - \beta_t' U_t$. Under the assumption that v_t and U_t are independent of each other and of both y_t and X_t , the error in y_t will be absorbed in the disturbance term of the regression and can be ignored while the errors in X_t will affect the estimation of β_t . If outliers affect a specific percentage of observations in the sample then classical measurement error and outliers result in an estimator that is both biased and inconsistent since $E[e_t^* | X_t^*] \neq 0$, $t = 1, \dots, T$. However, if the absolute number of additive outliers is fixed then $E[e_t^* | X_t^*]$ will converge to zero as $T \rightarrow \infty$.

Stock (1986) gave out a clear illustration of the asymptotical distributions for the $F_T(\frac{m}{T})$ -statistic in *SupF*, *AveF* and *ExpF* tests when the data is observed without error. Let “ \Rightarrow ” denote weak convergence in probability and “ \xrightarrow{P} ” convergence in probability. The asymptotic distribution of the F -test statistic is found to be $F_T(\lambda) \Rightarrow \frac{B_k^\mu(\lambda)' B_k^\mu(\lambda)}{\lambda(1-\lambda)}$, where

$\lambda = \frac{m}{T} \in (0,1)$, $B_k^\mu(\lambda) = W_k(\lambda) - \lambda W_k(1)$ and $W_k(\cdot)$ is a k -dimensional standard Brownian

motion. To show this result we first define the $k \times 1$ vector $v_T(\lambda) = T^{-1/2} \sum_{t=1}^{T\lambda} X_t e_t$ and the $k \times k$

matrix $V_T(\lambda) = T^{-1} \sum_{t=1}^{T\lambda} X_t X_t'$. By a Functional Central Limit Theorem we can get that

$v_T(\cdot) \Rightarrow \sigma_e \Sigma_X^{1/2} W_k(\cdot)$ and $V_T(\lambda) \xrightarrow{P} \lambda \Sigma_X$, where $\Sigma_X = E X_t X_t'$. Substituting the asymptotic distribution of $v_T(\cdot)$ and $V_T(\lambda)$ into the partial sum of squared residuals we obtain

$SSR_{1,T\lambda} = -(\sum_{t=1}^{T\lambda} X_{t-1} e_t)' (\sum_{t=1}^{T\lambda} X_t X_t')^{-1} (\sum_{t=1}^{T\lambda} X_t e_t) = -v_T(\lambda)' V_T(\lambda)^{-1} v_T(\lambda)$, and it follows that

$F_T(\lambda) = F_T\left(\frac{m}{T}\right) = \frac{SSR_{1,T} - (SSR_{1,m} + SSR_{m+1,T})}{(SSR_{1,m} + SSR_{m+1,T})/(T-2k)}$ converges to an asymptotic χ_k^2 distribution

under the null hypothesis.

When the data is observed with additive errors, however, the asymptotic distribution of the F statistic will be different since both $v_T(\lambda)$ and $V_T(\lambda)$ will be affected:

$$\begin{aligned} v^*(\lambda) &= T^{-1/2} \sum_{t=2}^{T\lambda} X_t^* e_t^* = T^{-1/2} \sum_{t=2}^{T\lambda} (X_t + U_t)(e_t + v_t - \beta_t' U_t) \\ &= T^{-1/2} \left(\sum_{t=2}^{T\lambda} X_t e_t - \sum_{t=2}^{T\lambda} U_t \beta_t' U_t \right) = v_T(\lambda) - T^{-1/2} \sum_{t=2}^{T\lambda} U_t \beta_t' U_t \\ V_T^*(\lambda) &= T^{-1} \sum_{t=2}^{T\lambda} X_t^* X_t^{*'} = T^{-1/2} \sum_{t=2}^{T\lambda} (X_t + U_t)(X_t + U_t)' = T^{-1} \left(\sum_{t=2}^{T\lambda} X_t X_t' + \sum_{t=2}^{T\lambda} U_t U_t' \right) \\ &= V_T(\lambda) + T^{-1} \sum_{t=2}^{T\lambda} U_t U_t' \end{aligned}$$

The additive contaminations in the dependent variable now disappear, making the $F_T(\lambda)$ statistic independent of the errors in the regressand. Since the perturbations in the regressor remain present, however, $v_T^*(\lambda)' V_T^*(\lambda)^{-1} v_T^*(\lambda)$ will be different from $v_T(\lambda)' V_T(\lambda)^{-1} v_T(\lambda)$. The

direction of the small sample bias of $F_T\left(\frac{m}{T}\right) = \frac{SSR_{1,T} - (SSR_{1,m} + SSR_{m+1,T})}{(SSR_{1,m} + SSR_{m+1,T})/(T-2k)}$ is hard to obtain

theoretically as both the denominator and the numerator of the test statistic are affected. We therefore perform a Monte Carlo experiment to investigate how the tests will be influenced with and without contaminations. The experiment design is similar to Kim and Perron (2009),

$$\begin{aligned} y_t &= \beta_t' X_t + u_t, u_t \sim i.i.d.N(0,1); \beta_t = \beta + \gamma * I(t > [T\lambda]) \\ X_t &= \mu + Z_t, Z_t = \rho Z_{t-1} + e_t, e_t \sim i.i.d.N(0,1) \end{aligned},$$

where

$$\mu = 1, \rho = 0.7, \gamma = \begin{cases} 0: & \text{for size investigation} \\ 0.4: & \text{for power investigation} \end{cases}, \beta = 0.4, \lambda = 0.5 \text{ and } T = 128, 256, 512^3.$$

We consider the affect of both types of error on the tests for structural change. Classical measurement error η_t is added in the regressor as $X_t^* = \mu + Z_t + \eta_t$, $\eta_t \in N(0, \sigma_\eta)$ and σ_η is determined by a signal to noise ratio (SNR) $SNR = SD(X)/\sigma_\eta$. $SNR = \infty$ indicates that the data is measured exactly whereas $SNR = 2.5$ and $SNR = 0.8$ represent imprecisely measured

³ The sample sizes are chosen to be suitable for the application of the wavelet transform (Percival and Walden 2000).

data with low and high error margins respectively. Table 1 shows the size and power of the three tests with and without measurement error based on 10,000 replications.

Table 1

The 95% confidence interval for actual size when the nominal size is 5% is given by $0.05 \pm 1.96 * \sqrt{\frac{0.05(1-0.05)}{10000}} = (0.0457, 0.0543)$ when there are 10,000 replications. Table 1 shows that for $SNR = 0.8$, the test size is distorted upwards while the powers is reduced. When $SNR = 2.5$ the test size remains mostly within the confidence interval for a true rejection rate of 5%, although a slight upwards tendency relative to the case without noise can be detected. The impact of measurement errors on test size does not, in general, diminish as the sample size grows. By contrast, the power becomes increasingly robust to measurement error as the number of observations increases. For a small magnitude of error (*i.e.* $SNR = 2.5$), a noticeable impact can only be found in the small sample case. The results in Table 1 indicate that the F -statistic based tests for structural change exhibit robustness to classical measurement error when its magnitude is not too excessive. The tests will, however, be influenced when the error margins around the true values are large. On the other hand, additive outliers (AO) in the explanatory variable are introduced by defining the distribution of η_t as a mixture between a point mass at zero and either positive or negative values from the Gumbel distribution:

$$\eta_t \sim \kappa \cdot \delta_0 + (1 - \kappa)(-1)^\pi \cdot Gumbel\left(\xi \cdot SD(X), \frac{1}{3}\right)$$

where $\pi \sim Bernoulli(0.5)$, $\kappa \in (0,1)$ and ξ specifies the outlier magnitude. In the simulation study we consider $\xi = 5$ and $\xi = 10$. Furthermore, we enforce a probability parameter $(1 - \kappa) = 0.015625$ by including $2^{\ln_2(T)-6}$ outliers at random positions in a sample of length T . Table 2 reports the size and power of the tests, again based on simulation with 10000 repetitions:

Table 2

Table 2 shows that size is greatly increased with the same direction of the magnitude of outliers. For the *SupF* test, these size distortions increase with the sample size. Similar tendencies cannot be found for the other tests except for the *AveF* test with large outliers, where size distortions decrease in larger samples. The power of tests for structural change decreases in the case with small outlier magnitude ($\xi = 5$) relative to the scenario without outliers. This tendency is reversed in small samples, leading to an increase in power when the magnitude outlier increases. In medium and large samples, however, the negative relation is maintained. All in all, test power become consistently more robust to outliers as the sample size increases. Among the three tests for structural change, the *SupF* and the *ExpF* statistics behave very similarly. The *AveF* statistic differs in that its size distortions are generally smaller than those of the other two tests. This does not, however, imply superior properties of the *AveF* test since its power losses in the presence of outliers generally are the largest among the three tests we have considered.

3. Improving the test performance using wavelet-based denoising

The traditional way to deal with measurement error in the linear model is to use an instrumental variable, which must be correlated with the true value of the regressors but not contemporaneously correlated with the measurement error. Appropriate instruments are, however, usually very difficult to find in practice (Cragg, 1994; Schenach, 2004; Gençay and Gradojevic, 2011). We therefore favour a wavelet based denoising approach to reduce the problems introduced by classical measurement errors. This method does not need any instrumental variables, is mostly data driven and allows the key features of a time series to be characterized by a small number of coefficients, thus facilitating the elimination of noise. A slightly modified wavelet-based denoising strategy performs remarkably well in detecting outliers, allowing their replacement with low frequency information from the time series. This procedure constitutes an alternative to regression-based models for outlier detection (*e.g.* Tsay, 1986) and will circumvent the risk of model misspecification.

Wavelet methods began to gain the attention of statisticians and econometricians after a series of articles in the field of economics and finance. Introductory texts for economists are given by Ramsey (1999), Schleicher (2002) and Crowley (2005) and more extensive descriptions have been provided by Vidakovic (1999), Percival and Walden (2000) and Gençay *et al.* (2001). Wavelet methodology represents an arbitrary time series in both time and frequency domains by convolution of the time series with a series of small wavelike functions..

Corresponding to the time infinite sinusoidal waves in the Fourier transform, the time located wavelet basis functions $\{\psi_{k,j} : k, j \in \mathbb{Z}\}$ used in the wavelet transform are generated by translations and dilations of a basic mother wavelet $\psi \in L^2(\mathbb{R})$. The function basis is constructed through $\psi_{k,j}(t) = 2^{j/2} \psi(2^j t - k)$, where k is the location index and j is the scale index which corresponds to the information inside the frequency band $(\frac{1}{2^j}, \frac{1}{2^{j-1}})$. For a signal f , its wavelet transform is the wavelet coefficients $f^* = \{\gamma(k, j)\}_{k,j \in \mathbb{Z}}$ with $\gamma(k, j) = \langle f, \psi_{k,j} \rangle = \int f(t) \psi_{k,j}^*(t) dt$, which can represent the resolution at time k and scale j . The resolutions in the time domain and the frequency domain can be achieved by shifting the time index k and the scale index j respectively. A lower level of j corresponds to higher frequency bands and higher level of j corresponds to lower frequency bands. Accordingly, the information at high frequency bands, such as noise, outliers, or data spikes, will be captured by $\gamma(k, j)$ at a lower level of j . By contrast, the long persistent information at low frequencies, *e.g.* trends or structural breaks, will be captured by $\gamma(k, j)$ at higher level of j . For a time series that is sampled at discrete time points, the coefficients of the time series in terms of the wavelet basis are obtained via the Discrete Wavelet Transform (*DWT*) and maximum overlap discrete wavelet transform (*MODWT*). The *DWT* can be implemented by applying a cascade of high-pass and low-pass filters to a time series that thereby separate its characteristics at different frequency bands (Mallat, 1989). For detailed illustration of *DWT* and *MODWT*, we refer to Vidakovic (1999), Percival and Walden (2000), and Gençay *et al.* (2001). Denoising of the time series is facilitated by three important properties of the wavelet transform: sparsity, orthogonality and linearity. Let $y = f + e$ be observed data with the unobservable signal f being contaminated by noise e . The *DWT* of y is defined as $d^* = Wy$, where W is a transformation matrix which combines high-pass and low-pass filters. Thus the transformation can filter out the information of the time series at different frequency levels. Linearity implies that d^* maintains the additive structure of y , *i.e.* $d^* = d + \varepsilon$ with $d = Wf$, $\varepsilon = We$. Orthogonality allows energy preservation so that $\|f\| = \|d\|$. Sparsity implies that the key characteristics of a systematic signal f , such as rapid local change, are captured in a small set of large *DWT* coefficients, leaving the remaining coefficients at a value of zero. By contrast, pure noise is spread evenly through all resolutions of d^* in the case of Gaussian white noise $\varepsilon_t \sim iidN(0, \sigma)$. In wavelet denoising, the *DWT* coefficients that are too

small are eliminated and we get the denoised coefficients \hat{d}^* . The denoised time-domain representation of the signal f can be reconstructed by applying the inverse wavelet transform on \hat{d}^* , i.e. $\hat{f} = W^T \hat{d}^*$. The choice of threshold is one of the most important steps and different approaches have been suggested in Donoho and Johnstone (1994, 1995) and Nason (1996). Since the sparsity of d is very important prior knowledge, we favour the empirical Bayes thresholding proposed by Johnstone and Silverman (2004, 2005a,b). These authors specify the prior distribution of the true signal as

$$f(d) = w\gamma(d) + (1-w)\delta_0(d),$$

where $0 \leq w \leq 1$ is the mixing weight, γ is some fat-tailed distribution and δ_0 is a point mass at 0. Based on the observed coefficients d^* and the distribution of ε , we can obtain the likelihood $f(d^*|d)$ and the posterior distribution $f(d|d^*)$. Finally, a Bayesian rule is set in order to choose a statistic \hat{d} as the estimator of d . This is usually the mean or median from the posterior distribution, where using the posterior median implies thresholding while using the posterior mean leads to wavelet shrinkage.

We will first apply empirical Bayes denoising to improve the results obtained in Table 1. In this case the terms of signal and error denote the true data X and the measurement error e . The heavy-tailed distribution in the prior is specified as a quasi-Cauchy distribution:

$$\gamma(d) = (2\pi)^{-1/2} \left\{ 1 - |u| \Phi(|d|) / \varphi(d) \right\}.$$

This corresponds to the assumption that the non-zero wavelet coefficients of the true data have a higher probability of being larger than a Gaussian component. If the measurement error e is distributed as $N(0, \sigma^2)$ the likelihood of the observed wavelet coefficients d^* will be $d^*|d \sim N(d, \sigma^2)$ and the cumulative posterior distribution of d is given by

$$F(d|d^*) = -(1 - e^{-d^{*2}/2})^{-1} \left\{ \Phi(d - d^*) - d^* \varphi(d - d^*) + (dd^* - 1)e^{dd^* - d^{*2}/2} \Phi(d) \right\}.$$

We judge the implied shrinkage rule to be more appropriate for the purpose of eliminating measurement error than thresholding, since complete elimination of small wavelet coefficients is more likely to remove minor details of the true signals together with the noise than a partial reduction of the coefficient values. Thus the posterior mean $u(d|d^*) = d^* (1 - e^{-d^{*2}/2})^{-1} - 2d^{*-1}$ is applied as the estimator for d . We conduct the same simulations as in Table 1, using denoised regressors for the breakpoint tests. Our results are shown in Table 3.

Table 3

It can be seen that denoising leads to small size corrections, which is most apparent at a SNR of 0.8. These downward corrections, however, appear in every scenario, indicating that the denoising algorithm modifies the data in a way that generally decreases size but does not specifically address the distortions caused by measurement error. The data modifications implied by the denoising method also affects the power. In small samples, a power decrease of up to 5% when there are no measurement errors has to be accepted. This negative influence decreases markedly with sample size, however, and the power differences in the case without measurement error are already less than one percentage point in samples with 256 observations. It can be seen that the denoising algorithm does not make a difference at a SNR of 2.5 when measurement error is added to the regressors. However, wavelet-based denoising helps to restore power in the scenario with more considerable error margins. The magnitude of these improvements amounts to around 12% in small and medium samples. We expect the overall performance of the denoising algorithm to improve when applied to time series with greater persistence. This is based on the fact that in more persistent time series the few DWT coefficients at coarser scales become larger relative to the greater number of coefficients at the finest scale. The amount of highly frequent data characteristics that are accidentally eliminated through the removal of measurement error will thus decrease.

The use of *empirical* Bayes thresholding for the removal of outliers reduces to identifying their position and focuses explicitly on the finest scale coefficients of the wavelet transform. It furthermore reverses the role of true data and perturbations seen previously. Since outliers concern solely one observation, most of their energy will be captured by wavelet coefficients at the finest scale. By contrast, as most economic time series experience some sort of persistence, their energy will for the most part be allocated in coarse scales, leaving only a small contribution at high frequencies. Applying empirical Bayes thresholding to the finest scale wavelet coefficients will therefore treat the details of the true time series as noise and isolate the outliers. Since the isolation of outliers requires complete elimination of all other data characteristics we conduct the denoising method using the threshold rule implied by the posterior median.

In order to correct the identified outliers, we first generate a benchmark series by applying the $MODWT$ on the contaminated series. We then eliminate all coefficients in the first four scales and reconstruct the data. This yields a series consisting solely of low frequency variation and

thus excludes the impact of outliers. In a next step, we define candidate points. Since the wavelet transform is localized, the position of influential observations can be determined from the non-zero coefficients of the denoised signal. At the highest scale, each wavelet coefficient is related to two observations in the underlying time series, yielding two candidates for the position of the outlier. We compare each pair of candidates to their values in the benchmark series that excludes high frequency variation. The observation that deviates most is replaced with the value from the benchmark series. This procedure has the advantage of circumventing the necessity of estimating further models that are prone to misspecification. Furthermore, the focus of the analysis is kept exclusively to short range patterns, thereby minimizing the risk of mistaking other characteristics of the data as outliers. Finally, the data correction is implemented as precisely as possible by modifying only the observation that has been identified as an outlier.

We conducted our simulation study as before, but this time using denoised regressors. The results are shown in Table 4:

Table 4

We can see that size reacts very favorably to this form of data correction. The excessive upward bias is reduced almost entirely, with only the exception of large outlier magnitudes in medium and large samples. In these cases an upward bias of 1.5% and 3% remains irrespective of which test is considered. The corrected data yields power results that are in very close to those obtained with the raw data when there are no outliers. The differences are in general smaller than 1.5% and indicate that the power of the tests for structural change is entirely maintained when using empirical Bayesian thresholding for outlier detection.

Despite being a powerful tool to clean the data from influential observations, outlier detection based on the thresholding procedure of Johnstone and Silverman (2004, 2005a,b) does not mistake characteristics of the data for outliers when there actually are none. Both size and power of the corrected data differ from the results obtained from the raw data by at most three decimal places in a scenario with an outlier magnitude of zero.

4. Conclusion

This paper first derives a mathematical formula to demonstrate that additive contaminations of the explanatory variables in a linear model will influence the widely used Quandt-Andrews

family of tests for structural breaks. Monte Carlo simulations show that classical Gaussian measurement error generally diminishes test power and inflates size when its magnitude is sufficiently large. Additive outliers are shown to cause excessive size distortion and reduced test power. We apply two denoising techniques, based on empirical Bayes thresholding, in order to remove perturbations from the data. In the case of measurement error, a wavelet shrinkage strategy fails to correct size, but succeeds in mitigating the negative impact on power when the error margin is large enough. An outlier removal procedure consisting of outlier detection via wavelet thresholding and outlier replacement with low-frequency information of the contaminated time series leads to very good results. The size and power of tests using the denoised data are maintained almost entirely.

References

- Abramovich, F., Sapatinas, T., and Silverman, B.W. (1998), "Wavelet thresholding via a Bayesian approach," *Journal of the Royal Statistical Society, Series B*, 60, 725-749.
- Andrews, D.W.K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 59, 817-858.
- Andrews, D.W.K. and Ploberger W. (1994), "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, 62, 1383-1414.
- Balke, N.S. and Fomby, T.B. (1994), "Large Shocks, Small Shocks, and Economic Fluctuations: Outliers in Macroeconomic Time Series", *Journal of Applied Econometric*, 9, 181-200.
- Bergeijk, P.A.G. (1995), "The accuracy of international economic observations", *Bulletin of Economic Research*, 47, 1-20.
- Bound, J., Brown, C., and Mathiowetz, N. (2001), "Measurement error in survey data", *Handbook of econometrics*, 5, 3705-3843.
- Brown, R.L., Durbin, J. and Evans, J.M. (1975), "Techniques for Testing the Constancy of Regression Relationships over Time", *Journal of the Royal Statistical Society, Series B*, 35, 149-192.
- Chow, G.C. (1960), "Tests of Equality between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 52, 211-222.
- Cragg, J.G. (1994), "Making good inferences from bad data", *Canadian Journal of Economics*, 27(4), 776-800.
- Crowley, P.M. (2005), "An Intuitive Guide to Wavelets for Economists", Bank of Finland Research. Discussion Papers No. 1.

- Davidson, R. and MacKinnon J. (2009), *Econometric Theory and Methods: International Edition*, Oxford University Press.
- Donoho, D.L. and Johnstone, I.M. (1994), “Ideal spatial adaptation by wavelet shrinkage”, *Biometrika*, 81, 425-455.
- Donoho, D.L. and Johnstone, I.M. (1995), “Adapting to unknown smoothness via wavelet shrinkage”, *Journal of the American Statistical Association*, 90, 1200-1224.
- Gençay, R. and Gradojevic, N. (2011), “Errors-in-variables estimation with wavelets”, *Journal of Statistical Computation and Simulation*, 81(11), 1545-1564.
- Gençay, R., Selçuk, F. and Whitcher, B. (2001), *An Introduction to Wavelets and Other Filtering Methods in Finance and Economic*, Academic Press, San Diego, CA, USA.
- Grossman, A. and Morlet, J. (1984), “Decomposition of Hardy functions into square integrable wavelets of constant shape”, *Society for Industrial and Applied Mathematics Journal on Mathematical Analysis*, 15, 732-736.
- Kim, D. and Perron, P. (2009), “Assessing the Relative Power of Structural Break Tests Using a Framework Based on the Approximate Bahadur Slope”, *Journal of Econometrics*, 149, 26-51.
- Kuan, C. and Hornik, K. (1995), “The generalized fluctuation test: A unifying view”, *Econometric Reviews*, 14(2), 135-161.
- Lucas, R.E. (1976), “Econometric policy evaluation: A critique”, *Carnegie-Rochester Conference Series on Public Policy*, 1, 19-46.
- Mallat, S.G. (1989), “Multiresolution approximation and wavelets orthonormal bases of $L^2(R)$ ”, *Transactions of the American Mathematical Society*, 315(1), 69-88.
- Morgenstern, O. (1963), *On the accuracy of economic observations*, 2nd Ed., Princeton University Press.
- Nason, G.P. (1996), “Wavelet shrinkage using cross-validation”, *Journal of the Royal Statistical society, Series B*, 58, 463-479.
- Hansen, B.E. (1992b), “Testing for Parameter Instability in Linear Models”, *Journal of Policy Modeling*, 14(4), 517-533.
- Johnstone, I.M. and Silverman, B.W. (2004), “Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences”, *Annals of Statistics*, 32, 1594-1649.
- Johnstone, I.M. and Silverman, B.W. (2005a), “Ebayesthresh: R programs for empirical Bayes thresholding,” *Journal of Statistic Software*, 12, 1-38.
- Johnstone, I.M. and Silverman, B.W. (2005b), “Empirical Bayes selection of wavelet thresholds,” *Annals of Statistics*, 33, 1700-1752.

Percival, D.B. and Walden, A.T. (2000), *Wavelet Methods for Time Series Analysis*, Cambridge Univ. Press.

Quandt, R.E. (1960), "Tests of Hypotheses that a Linear System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324-330.

Ramsey, J. (1999), "The contribution of wavelets to the analysis of economic and financial data," *Phil. Trans. R. Soc. Lond. A* 357: 2593-2606.

Rodrigues, P.M. and Rubia, A. (2011), "The effects of Additive Outliers and Measurement Errors when Testing for Structural Breaks in Variance", *Oxford Bulletin of Economics and Statistics*, 73(4), 449-468.

Schennach, S.M. (2004), "Exponential specifications and measurement error," *Economic Letters*, 85, 85-91.

Schleicher, C. (2002), "An Introduction to Wavelets for Economists", Monetary and Financial Analysis Department, Bank of Canada, Working paper 2002-2003.

Stock, J.H. (1994), Unit roots, structural breaks and trends. *Handbook of econometrics*, 4, 2739-2841.

Tsay, R.S. (1986), "Time Series Model Specification in the Presence of Outliers", *Journal of the American Statistical Association*, 81, 132-141.

Vidakovic, B. (1999), *Statistical modeling by wavelets*, Vol. 503, John Wiley & Sons.

Table 1: Size and power of tests using raw data with measurement error

| <i>Test</i> | <i>#obs</i> | <i>Signal-to-noise ratio</i> | | |
|------------------------|-------------|------------------------------|--------|--------|
| $\gamma = 0$ (size) | | ∞ | 2.5 | 0.8 |
| $SupF_T$ | 128 | 0.0454 | 0.0463 | 0.0811 |
| | 256 | 0.0456 | 0.0482 | 0.0939 |
| | 512 | 0.0494 | 0.0509 | 0.0959 |
| $AveF_T$ | 128 | 0.0501 | 0.052 | 0.0896 |
| | 256 | 0.0476 | 0.0543 | 0.0886 |
| | 512 | 0.049 | 0.0506 | 0.0858 |
| $ExpF_T$ | 128 | 0.0523 | 0.056 | 0.0979 |
| | 256 | 0.0503 | 0.0532 | 0.0991 |
| | 512 | 0.0494 | 0.0537 | 0.0943 |
| $\gamma = 0.4$ (power) | | | | |
| $SupF_T$ | 128 | 0.8725 | 0.7991 | 0.461 |
| | 256 | 0.9951 | 0.9844 | 0.7295 |
| | 512 | 1 | 1 | 0.9424 |
| $AveF_T$ | 128 | 0.9051 | 0.845 | 0.5079 |
| | 256 | 0.9962 | 0.987 | 0.7527 |
| | 512 | 1 | 1 | 0.9467 |
| $ExpF_T$ | 128 | 0.9064 | 0.8463 | 0.5156 |
| | 256 | 0.997 | 0.9887 | 0.7585 |
| | 512 | 1 | 1 | 0.9509 |

Table 2: Size and power of tests using raw data with outliers

| <i>Test</i> | <i># obs</i> | <i>Outlier magnitude</i> | | |
|------------------------|--------------|--------------------------|--------|--------|
| $\gamma = 0$ (size) | | 0 | 5 | 10 |
| $SupF_T$ | 128 | 0.0432 | 0.2304 | 0.6804 |
| | 256 | 0.0448 | 0.2709 | 0.7354 |
| | 512 | 0.0488 | 0.2910 | 0.7430 |
| $AveF_T$ | 128 | 0.0496 | 0.2149 | 0.6289 |
| | 256 | 0.0469 | 0.2231 | 0.6006 |
| | 512 | 0.0526 | 0.2130 | 0.5573 |
| $ExpF_T$ | 128 | 0.051 | 0.2527 | 0.7048 |
| | 256 | 0.0497 | 0.2702 | 0.7203 |
| | 512 | 0.0527 | 0.2695 | 0.7029 |
| $\gamma = 0.4$ (power) | | | | |
| $SupF_T$ | 128 | 0.8641 | 0.7841 | 0.8878 |
| | 256 | 0.9948 | 0.9214 | 0.9161 |
| | 512 | 1.0000 | 0.9911 | 0.9549 |
| $AveF_T$ | 128 | 0.8978 | 0.7359 | 0.8181 |
| | 256 | 0.9962 | 0.8971 | 0.8350 |
| | 512 | 1.0000 | 0.9871 | 0.8933 |
| $ExpF_T$ | 128 | 0.8980 | 0.7979 | 0.8914 |
| | 256 | 0.9971 | 0.9216 | 0.9075 |
| | 512 | 1.0000 | 0.9913 | 0.9443 |

Table 3: Size and power of tests using the denoised data in the case of measurement error

| <i>Test</i> | <i>#obs</i> | <i>Signal-to-noise ratio</i> | | |
|------------------------|-------------|------------------------------|--------|--------|
| $\gamma = 0$ (size) | | | 2.5 | 0.8 |
| $SupF_T$ | 128 | 0.0426 | 0.0431 | 0.0839 |
| | 256 | 0.0391 | 0.043 | 0.089 |
| | 512 | 0.0448 | 0.0476 | 0.0895 |
| $AveF_T$ | 128 | 0.0473 | 0.0476 | 0.0817 |
| | 256 | 0.0458 | 0.0516 | 0.0845 |
| | 512 | 0.0439 | 0.0484 | 0.0756 |
| $ExpF_T$ | 128 | 0.0489 | 0.0492 | 0.0912 |
| | 256 | 0.0448 | 0.0489 | 0.0918 |
| | 512 | 0.0461 | 0.0474 | 0.0857 |
| $\gamma = 0.4$ (power) | | | | |
| $SupF_T$ | 128 | 0.8225 | 0.7682 | 0.5833 |
| | 256 | 0.9895 | 0.9753 | 0.8543 |
| | 512 | 1 | 1 | 0.9866 |
| $AveF_T$ | 128 | 0.8666 | 0.8216 | 0.6244 |
| | 256 | 0.9908 | 0.9813 | 0.8657 |
| | 512 | 1 | 0.9999 | 0.9877 |
| $ExpF_T$ | 128 | 0.8676 | 0.8208 | 0.6309 |
| | 256 | 0.9922 | 0.983 | 0.876 |
| | 512 | 1 | 1 | 0.9891 |

Table 4: Size and power of tests using the denoised data in the case of additive outliers

| <i>Test</i> | <i># obs</i> | <i>Outlier magnitude</i> | | |
|------------------------|--------------|--------------------------|--------|--------|
| $\gamma = 0$ (size) | | 0 | 5 | 10 |
| $SupF_T$ | 128 | 0.0431 | 0.0436 | 0.0488 |
| | 256 | 0.0447 | 0.0487 | 0.0633 |
| | 512 | 0.0487 | 0.0545 | 0.0866 |
| $AveF_T$ | 128 | 0.0492 | 0.0506 | 0.0562 |
| | 256 | 0.0468 | 0.0521 | 0.0662 |
| | 512 | 0.0523 | 0.0533 | 0.085 |
| $ExpF_T$ | 128 | 0.0493 | 0.0497 | 0.056 |
| | 256 | 0.049 | 0.0527 | 0.0667 |
| | 512 | 0.0522 | 0.0564 | 0.0882 |
| $\gamma = 0.4$ (power) | | | | |
| $SupF_T$ | 128 | 0.8593 | 0.8508 | 0.8522 |
| | 256 | 0.9943 | 0.9934 | 0.9937 |
| | 512 | 1 | 1 | 1 |
| $AveF_T$ | 128 | 0.8944 | 0.887 | 0.8882 |
| | 256 | 0.9959 | 0.9944 | 0.9942 |
| | 512 | 1 | 1 | 0.9996 |
| $ExpF_T$ | 128 | 0.8952 | 0.8888 | 0.8901 |
| | 256 | 0.997 | 0.9956 | 0.9962 |
| | 512 | 1 | 1 | 1 |