LUNDS UNIVERSITET
Ekonomihögskolan

Are tests for smooth structural change affected by data inaccuracies?

a simulation study

August 31, 2012

Abstract

The size and power of tests for smooth structural change are evaluated in the presence of random measurement error in the explanatory variable or outliers in the dependent variable of a univariate regression model. It is shown that the considered tests are robust to measurement error of a magnitude that can be found in real economic data. By contrast, outliers are found to distort both the size and the power of test for structural breaks. It is shown that the effects of outliers can be compensated by a simple wavelet-based outlier detection algorithm.

Keywords: Structural breaks, measurement error, additive outliers, wavelet analysis

NEKN01 - master essay I
Author: Simon Reese, eco11sre@student.lu.se
Supervisors: Yushu Li, Fredrik NG Andersson

Contents

List of Tables

1 Introduction

The detection of presumed structural breaks is an integral part of economic time series analysis. If an econometric model is meant 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). Still, studies of longitudinal data are often conducted under the assumption of parameter constancy over the entire sample length. But ignoring the presence of a structural break in the sample generally leads to a misspecification of the econometric model and can have severe consequences for the practical use of the analysis. More specifically, it is difficult to rely on statistical inference and forecasts from the model in this situation (see for example Stock and Watson, 2007, p. 565).

This thesis attempts to measure the size and power distortions of tests for smooth structural changes in the presence of data disturbances using Monte Carlo simulations. More explicitly, tests that show power against gradual parameter shifts are evaluated in the presence of either measurement error in the explanatory variable or outliers in the dependent variable. It is observed that the considered testing procedures are robust to measurement error, but that outliers constitute a severe threat to size and power of all tests, irrespective of the chosen scenario. In order to correct the data, a wavelet-based outlier detection algorithm is applied. Further simulations show that this correction method helps to mitigate size and power distortions to a large extent.

Procedures to identify the existence of structural change in an econometric model have attracted substantial research effort. The most influential contribution in this regard is Chow's (1960) analysis-of-variance test which is a comparison of the sum of squared residuals in models with and without parameter change. An extension of the testing principle to situations where the break point is unknown has been proposed by Quandt (1958, 1960), who suggests taking the highest test statistic in within the set of Chow test statistics at all possible break points, a method that is know as the Quandt likelihood ratio test. Critical values for this test have been provided by Andrews (1993). Andrews and Ploberger (1994) have suggested an optimal alternative test to Quandt's procedure. Instead of choosing the maximum test statistic over all possible break points, the authors calculate the average exponential test statistic. They show further that another principle, namely the simple average of all test statistics, can be derived as a special case of their optimal method. Not requiring the break date to be known ex ante allows to apply the tests in a vast variety of situations and according to Hansen (2001, p.121) "[t]he Quandt-Andrews and Andrews-Ploberger family of statistics have essentially replaced the Chow statistic in recent econometric practice." Applications of the tests can be found in e.g. McConnell and Perez-Quiros (2000), Dees et al. (2007) or Laubach (2009).

An alternative testing approach is the CUSUM test of Brown et al. (1975) that investigates the cumulated recursive residuals of a model. Its is designed to capture more general forms of structural change, but suffers from a tendency to lose power for structural change that occurs late in the sample. Bauer and Hackl (1978) propose MOSUM, a moving average of the recursive residuals, to improve on this problem of the CUSUM test.

In many situations it is plausible to assume that structural change does not happen immediately, but develops over a certain period. This reasoning can be motivated by the inability of economic agents to adjust immediately to a changing environment. Yet, structural breaks are usually assumed to be discrete the sake of parsimony and conceptual simplicity, even though they might not be intuitively plausible (Hansen, 2001, p. 118). Testing procedures for discrete breaks may have power against smooth structural change. But as shown in Chen and Hong (2008), their power is outperformed by testing procedures that directly address gradual parameter changes. In contrast to the extensive literature on abrupt structural breaks, only few articles focus on testing for smooth structural change. Farley et al. (1975) provide a first modeling approach by assuming the time-varying parameter to be a continuous linear function of time. Lin and Teräsvirta (1994) expand this approach to non-linear shifts by proposing a smooth transition function (STR) depending on time. Depending on the chosen function, the STR method allows to address monotonous as well as non-monotonous smooth changes and to establish the most powerful test for a known transition form in the sample. However, Farley et al.'s and Lin and Teräsvirta's models have the disadvantage of imposing restrictive ex-ante assumptions about the exact nature of structural change and of testing the model only against a very specific chosen alternative. By contrast, Chen and Hong (2012a) circumvent the need to specify the nature of the parameter shift by using non-parametric methods, or more specifically local linear smoothing. Their approach does not involve any assumption about the parameter vector β. Instead, the authors estimate the value of β at every point in the sample by conducting linear regressions in a local subsample around the respective time point. Their unrestricted parameter estimator is obtained by gathering all values of β from the local regressions. This approach leads potentially to great flexibility with regards to different smooth shifts that may appear in the data. Using the non-parametric parameter estimator, the null hypothesis of no structural change is testing by applying either a general Chow test or a general Hausman test.

With regards to the intention of this thesis to measure the size and power distortions of tests for smooth structural changes in the presence of data disturbances, the only study that is related to this idea to some extent is performed by Li (2012). The author examines the effect of measurement error in a setup with discrete structural changes and shows that the tests of Andrews (1993) and Andrews and Ploberger (1994) experience significant power losses when the explanatory variable of a univariate model is measured with error.

Besides targeting smooth structural change, this thesis extends the perspective with respect to relevant testing procedures. The tests in question are the Quandt LR test plus the average Chow and exponential average Chow procedures of Andrews and Ploberger (1994) on the one hand as well as the two tests for smooth structural break of Chen and Hong (2012a) on the other hand. Their performance will be evaluated given two different types disturbances in the data: First, measurement error in the explanatory variable of an econometric model is taken into account. Such error margins are a quite common problem in empirical work, as economic statistics practically cannot be produced without some degree of imperfection or incompleteness (Morgenstern, 1963, p. 13). Additionally, the consequences of outliers in the dependent variable are considered. As shown by Balke and Fomby (1994), their presence is given in most macroeconomic indicators and accounts for a significant share of the overall variance in the sampling period. Finally, the scenario choice of the analysis follows Chen and Hong (2008) by putting emphasis on either a monotonous smooth shift, a non-monotonous equivalent or a series of highly frequent smooth shifts.

The results show in a first instance that observations of Li (2012) cannot be made in the case of smooth structural change. In fact, all considered tests prove to be robust to measurement error. This fact is, however, strongly related to the chosen, empirically motivated magnitude of measurement error in the study and it may be guessed that substantially higher error margins will have a significant detrimental effect. In a second instance, it is shown that outliers in the dependent variable lead to an increased probability of Type 1 errors for most test in the study, although the dimension of the changes is not excessive. Additionally, it is observed that the power of tests for smooth structural breaks is substantially altered by the presence of outliers, where the extent of power losses depends on the outlier magnitude and the relative amount of outliers in the sample. A wavelet-based algorithm, suggested by Grané and Veiga (2010), is then applied to detect and eliminate the outliers. Conducting a simulation study with the corrected data, it is shown that the power losses due to outliers in the data can be compensated to a large extent.

Section 2 contains an exposition of the tests that are used in this thesis. The performance of tests for smooth structural breaks in the presence of measurement error and outliers is evaluated using a simulation study in Section 3. Section 4 introduces the algorithm of Grané and

Veiga (2010) and repeats the simulation study for the corrected data. Finally, section 5 concludes.

2 Testing for smooth structural change

Andrews (1993) and Andrews and Ploberger (1994) propose three different test statistics that are applied in situations where the timing of a discrete break point is unknown. The implementation of the tests can employ either a Wald, a Lagrange multiplier or a likelihood ratio test, where all three variations are asymptotically equivalent. Emphasis will be put on the Wald form of the testing statistics, as implemented by Zeileis et al. (2002) for the program package **R**.

Due to the absence of the break point parameter in the null hypothesis of tests for structural breaks at unknown time point, the asymptotic distribution of the related test statistic has initially been unknown and was first derived by Kim and Siegmund (1989). Andrews (1993) presents a fully operational version, including asymptotic critical values, of the Quandt likelihood ratio (Quandt, 1958, 1960) test that consists in picking the largest value from breakpoint tests at every point π in an interval of possible breakpoints $\Pi = [\underline{\pi}; \overline{\pi}]$. The test statistic is hence given by

$$\textit{sup-F} = \sup_{\pi \in \Pi} F(\pi) \tag{2.1}$$

Two additional testing principles for Wald, LM and LR-like tests are suggested by Andrews and Ploberger (1994). The authors introduce a test statistic that takes the (weighted) average of the exponentials of the different test statistics at each presumed break data. Its simple form from Zeileis et al. (2002) is given by

$$\textit{exp-F} = \log\left(\frac{1}{\overline{\pi} - \underline{\pi} + 1} \sum_{i=\underline{\pi}}^{\overline{\pi}} \exp(0.5 F_i)\right) \tag{2.2}$$

Andrews and Ploberger (1994) show further that a special case of their test is a simple average over all individual breakpoint tests:

$$\textit{ave-F} = \frac{1}{\overline{\pi} - \underline{\pi} + 1} \sum_{i=\underline{\pi}}^{\overline{\pi}} F_i \tag{2.3}$$

As in the case of the sup-F test, deriving the asymptotic distribution of the test statistic is a fairly difficult issue due to the fact that the formulation of the hypotheses is nonstandard. However, Andrews and Ploberger (1994) provide asymptotic critical values.

The three tests of Andrews (1993) and Andrews and Ploberger (1994), henceforth denoted Andrews-Ploberger tests, are initially tests for discrete structural breaks. Still, Andrews (1993) points out the potential of the Quandt LR test in detecting more general break forms. This potential has indeed been observed by Chen and Hong (2008, 2012a) for all three Andrews-Ploberger tests in the case of smooth structural change. Chen and Hong also propose two new tests that are specifically designed to identify gradual parameter shifts. They employ local linear smoothing, a non-parametric estimation technique to avoid making assumptions about the form of the transmission. Instead, the coefficient vector under the alternative is estimated by a collection of parameter values from local regressions. This allows to test for parameter constancy under the alternative of having a varying coefficient vector that is some undefined function of time, i.e.

$$H_0: \ \beta_t = \beta_0 \ \forall\, t \qquad H_a: \ \beta_t = \beta_a\left(\frac{t}{T}\right)$$

The time-varying parameter under the alternative is obtained by running local regressions around every time point in the sample. These regressions include the initial regressors and interactions of the explanatory variables with the temporal distance to the midpoint of the local regression relative to the entire sample size. The model around a point t is formally denoted as

$$Y_s = \alpha'_{t,0}\mathbf{X}_s + \alpha'_{t,1}\left(\frac{s-t}{T}\right)\mathbf{X}_s \tag{2.4}$$

The subsample is chosen according to a bandwidth $h = \frac{1}{\sqrt{12}}T^{-0.2}$, a value that Chen and Hong assume for convenience. This implies that the $\lfloor hT \rfloor$ observations before and after the time point of interest are included in the sample:

$$s \in [t - \lfloor hT \rfloor; t + \lfloor hT \rfloor]$$

where $\lfloor hT \rfloor$ is the next integer below hT. In order to allow the estimation of the local model at the beginning and the end of the sample, Chen and Hong create pseudo-data by reflecting the sample at its borders.

Moreover, the local parameters are estimated with weighted least squares, using a kernel function with respect to $\frac{s-t}{hT}$. That is, observations that are far from the center of the subsample contribute only little to the estimation of the regression coefficients. Having obtained the WLS estimators of the local regression around t, the value of the regression coefficients of the global model at time t is the value of the local coefficient vector $\alpha_{t,0}$. By contrast, the the local interaction coefficients between the regressors and time are not used in the global model. Finally, the global non-parametric estimator $\hat{\beta}_a\left(\frac{t}{T}\right),\ t = 1, \ldots, T$ is obtained by collecting the estimated coefficients $\hat{\alpha}_{t,0}$ from all local regressions.

Using this time-varying estimator, Chen and Hong (2012a,b) propose two tests:

1. A generalized Chow test. The test statistic is given by

$$\hat{C} = \frac{\sqrt{h}(SSR_0 - SSR_a) - \hat{A}_C}{\sqrt{\hat{B}_C}} \tag{2.5}$$

   where SSR_0 and SSR_a denote the sum of squared residuals for a model with standard OLS parameters and a model with the variable non-parametric coefficients respectively. Furthermore,

$$\hat{A}_C = h^{-\frac{1}{2}} dC_{AC}\hat{\sigma}^2 \quad \text{and} \quad \hat{B}_C = 4dC_{BC}\,\hat{\sigma}^4 \tag{2.6}$$

   with

$$C_{AC} = 2k(0) + h - \frac{1}{Th}\sum_{j=-\lfloor Th \rfloor}^{\lfloor Th \rfloor}\left(1 - \frac{|j|}{T}\right)k\left(\frac{j}{Th}\right)\left[k\left(\frac{j}{Th}\right) + h\int_{-1}^{1}k\left(\frac{j}{Th} + 2u\right)du\right] \tag{2.7}$$

$$C_{BC} = \frac{1}{Th}\sum_{j=1}^{T-1}\left(1 - \frac{j}{T}\right)\left[2k\left(\frac{j}{Th}\right) - \int_{-1}^{1}k(u)k\left(u + \frac{j}{Th}\right)du\right]^2 \tag{2.8}$$

   are centering and scaling parameters that normalize the asymptotic distribution of the test statistic to $N(0,1)$. As in the classical Chow test, the test statistic measures whether a model that allows for variable coefficients is able to explain the data better than a model that assumes constant parameters.

2. A generalized Hausman test. Its test statistic is denoted as

$$\hat{H} = \frac{T\sqrt{h}\hat{Q} - \hat{A}_H}{\sqrt{\hat{B}_H}} \tag{2.9}$$

where

$$\hat{Q} = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{X}'_t\hat{\beta}_a\left(\frac{t}{T}\right) - \mathbf{X}'_t\hat{\beta}_0\right)^2 \tag{2.10}$$

Moreover,

$$\hat{A}_H = h^{-\frac{1}{2}} dC_{AH}\hat{\sigma}^2 \quad \text{and} \quad \hat{B}_H = 4dC_{BH}\,\hat{\sigma}^4 \tag{2.11}$$

with

$$C_{AH} = \frac{1}{Th}\sum_{j=-\lfloor Th\rfloor}^{\lfloor Th\rfloor}\left(1 - \frac{|j|}{T}\right)k\left(\frac{j}{Th}\right)\left[k\left(\frac{j}{Th}\right) + h\int_{-1}^{1}k\left(\frac{j}{Th} + 2u\right)du\right] \tag{2.12}$$

$$C_{BH} = \frac{1}{Th}\sum_{j=1}^{T-1}\left(1 - \frac{j}{T}\right)\left[\int_{-1}^{1}k(u)k\left(u + \frac{j}{Th}\right)du\right]^2 \tag{2.13}$$

normalize the asymptotic distribution of the test statistic to a standard normal distribution. The general intuition behind the test is to look, whether the fits of the model with non-parametric estimator differ significantly from those of a simple OLS model. If so, the parameter vector must vary somehow with respect to time, which implies a smooth structural change.

Chen and Hong also provide a heteroskedasticity-robust version of the centering and scaling parameters. The Chow centering and scaling parameters are reformulated

$$\hat{A}_{C-het} = h^{-\frac{1}{2}}C_{AC}\,\text{trace}(\hat{\Omega}\hat{M}^{-1}) \quad \text{and} \quad \hat{B}_{C-het} = 4C_{BC}\,\text{trace}(\hat{M}^{-1}\hat{\Omega}\hat{M}^{-1}\hat{\Omega}) \tag{2.14}$$

where $\hat{M} = T^{-1}\sum_{t=1}^{T-1}\mathbf{X}_t\mathbf{X}'_t$ and $\hat{\Omega} = T^{-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_t^2\mathbf{X}_t\mathbf{X}'_t$. Likewise, the corresponding parameters in the Hausman test are modified to

$$\hat{A}_{H-het} = h^{-\frac{1}{2}}C_{AH}\,\text{trace}(\hat{\Omega}\hat{M}^{-1}) \quad \text{and} \quad \hat{B}_{H-het} = 4C_{BH}\,\text{trace}(\hat{M}^{-1}\hat{\Omega}\hat{M}^{-1}\hat{\Omega}) \tag{2.15}$$

As the tests procedures of Andrews (1993), Andrews and Ploberger (1994) and Chen and Hong (2012a) involve estimating a large number of regressions, it can be suspected that they are sensitive to inaccuracy in the data. With regards to measurement error, the resulting attenuation bias of the estimated coefficients (see for example Bound et al., 2001, p. 3712ff) provides a rationale to suspect an influence on test powers and sizes in general.

As regards outliers, the bias they may introduce into the estimated coefficients via the weight that is assigned to influential observations in OLS estimation may be specifically harmful if the sample is split up in two sub-samples: The coefficients in both samples are likely to differ solely because of the presence of an outlier in one sub-sample. Consequently, it can be assumed that tests for structural breaks tend to overreject, as local regression models are part of all considered testing procedures.

3 Monte Carlo simulations

The impact of measurement errors and outliers on the power of tests for smooth structural change is studied using **R**, as the program provides the required routines for the Andrews-Ploberger tests and especially wavelet methods via user-written commands. The simulation setup is oriented towards the model specification of Chen and Hong (2008, 2012a). The authors design a simple univariate regression framework where the explanatory variable itself is modeled as a stationary AR(1)-process. Formally, this is expressed as:

$$Y_t = F(\tau)(1 + 0.5X_t) + \varepsilon_t$$
$$X_t = 0.5X_{t-1} + \nu_t$$
$$\varepsilon \sim i.i.d.(0,1) \text{ and } \nu \sim i.i.d.(0,1)$$

Smooth structural change is introduced into the model through the function $F(\tau)$, which depends on the relative position of the observation at time point t within the entire sample. In the case of no structural change, $F(\tau)$ can be set to one. Apart from this simple case, three scenarios from Chen and Hong (2008) will be considered:

1. A monotonous smooth structural change:

$$F(\tau) = \left(1 + \exp[-1500(\tau^3 - 3\tau^2 + 3\tau - 1)]\right)^{-1}$$

2. A non-monotonous smooth structural change:

$$F(\tau) = 1 - \exp[-3(\tau - 0.5)^2]$$

3. A smooth structural change with high frequency:

$$F(\tau) = 0.5 + \sin(25\pi\tau)$$

For the case of measurement error, it will be assumed that the true explanatory variable X_t cannot be observed exactly. Instead, the information at disposal consists of the observed variable $X_t^* = X_t + \xi_t,\ \xi_t \sim i.i.d(0, \sigma_\xi)$, which expresses the erroneously measured data for the explanatory variable. Information about the variance of the measurement error can be translated to match the notion of signal and noise from wavelet thresholding. The signal-to-noise ratio will be used, which is the ratio of the sum of squared deviations from zero between the true time series and the measurement error. Formally, it is given by

$$SNR = \frac{\sum_{t=1}^{T} X_t^2}{\sum_{t=1}^{T} \xi_t^2} \tag{3.1}$$

In the case at hand, these sums of squared deviations from zero are equal to the variances of $\mathbf{X}$ and ξ. In order to make the concept more tangible in the case of no measurement error, the *inverse* signal-to-noise ratio will be considered (which will be called the noise-to-signal ratio NSR). Reasonable values for the NSR can be derived from previous studies that attempted to measure the extent of measurement error in economic data. These exercises are in most cases conducted for international trade data, where an exemplary case is the calculations of van Bergeijk (1995) and Makhoul and Otterstrom (1998) who indicate an error margin of around 8.5%-10% for OECD countries and 29% for non-OECD countries. Likewise, confidence intervals in the quarterly labor market reports of Statistics Sweden lead to error margins that vary mostly between 2% and 15%, depending on the extent of aggregation. The model setup is oriented towards an upper limit of these values by assuming an error margin of 15% and 30% respectively. In the latter case, this implies that the mean absolute deviation (MAD) of the measurement error from zero is 0.3 times the MAD of the true time series from zero. Taking the square of this number results in the respective target NSR of 0.09. In the following simulation this value will be chosen as well as a smaller value of 0.0225, the NSR-equivalent to an error margin of 15%. Finally, the choice of sample size deviates slightly from the values that Chen and Hong (2012a) choose, as dyadic sample sizes are chosen to prepare the use of wavelet methods for data correction. Fortunately, the closest powers of two to the sample sizes of 100, 250 and 500 are given by the numbers 128, 256 and 512, which implies only very slight changes. The simulation runs with 1000 iterations and incorporates the F-sup, F-ave and F-exp tests of Andrews (1993) and Andrews and Ploberger (1994) plus the generalized Chow and Hausman tests of Chen and Hong (2012a) in their homoskedasticity-specific and their heteroskedasticity-robust form. Information about the size of the tests is collected and their power is measured for all possible combinations of model type, sample size and noise-to-signal ratio. All results are based on empirical critical values at a significance level of 95%.

The effects of outliers are studied in the same model setup that was given for measurement error. But instead of altering the explanatory variable, adjustments will be made to the dependent variable. Emphasis is put on type 1 or additive outliers, i.e. an outlier that affects the observation at the same time point without having any influence on subsequent observations (Fox, 1972, p. 351). At four random time points in the sample, an impulse of $\Delta = z * \sigma_y$ will be added to the observed value, where z is an arbitrarily chosen magnitude parameter. The two cases $z_1 = 5$ and $z_2 = 10$ will be considered. Define a set $T^* = \{t_1^*, \ldots, t_4^*\}$ that contains the position of the four outliers. The model including outliers is then given by

$$Y_t = \begin{cases} F(\tau)(1 + 0.5X_t) + \varepsilon_t & \text{if } t \notin T^* \\ F(\tau)(1 + 0.5X_t) + \Delta + \varepsilon_t & \text{if } t \in T^* \end{cases} \tag{3.2}$$

where the properties of $F(\tau)$, X_t, and ε_t remain unchanged.

Table 1 depicts the resulting test sizes for the chosen scenarios. It becomes obvious that measurement error does not have a significant influence on the size of test for smooth structural change. The obtained values remain more or less constant, regardless of how large the error margin is for the explanatory variable. Solely for Chen and Hong's Hausman-type test, a marginal increase can be observed, although the change of at most 1.6 percentage points lacks practical relevance. However, the exact value do not remain completely unchanged. This suggests some form of impact of measurement error that is negligible in the chosen setup. With regards to outliers, it can be observed that their presence generally leads to inflated size when the number of observations is small. A clear positive relation between outlier magnitude and test sizes can be observed in this case. In medium-sized and large samples, size distortions are not given for the smallest chosen outlier magnitude, but appear when outliers have a dimension of ten standard deviations of the dependent variable. In addition, a negative relation between the sample size and the test size increases can be observed, which is at least partially due to the relative amount of outliers within the entire sample. It is worth noting, that the test of Chen and Hong react on average stronger than the Andrews-Ploberger tests and notably the Ave-F test, whose size remains approximately constant over all considered outlier magnitudes. An exception to this observation is the heteroskedasticity-robust version of Chen and Hong's (2012a) Hausman-type test, which experiences only very slight changes. All other tests have size increases of up to ten percentage points when outliers are present.

Table 1: Test sizes

# obs	128			256			512		
nts ratio	*0*	*0.15*	*0.3*	*0*	*0.15*	*0.3*	*0*	*0.15*	*0.3*
test									
Sup-F	0.043	0.047	0.051	0.058	0.044	0.043	0.051	0.045	0.045
Ave-F	0.036	0.052	0.056	0.051	0.053	0.053	0.057	0.059	0.057
Exp-F	0.041	0.053	0.054	0.058	0.048	0.045	0.047	0.052	0.043
C	0.028	0.023	0.025	0.036	0.031	0.033	0.034	0.038	0.039
C_{het}	0.043	0.041	0.047	0.051	0.051	0.05	0.046	0.049	0.051
H	0.155	0.143	0.147	0.126	0.119	0.12	0.097	0.107	0.101
H_{het}	0.078	0.072	0.073	0.082	0.069	0.067	0.066	0.059	0.067
Outlier magnitude	*0*	*5*	*10*	*0*	*5*	*10*	*0*	*5*	*10*
test									
Sup-F	0.043	0.072	0.117	0.058	0.049	0.075	0.051	0.051	0.063
Ave-F	0.036	0.055	0.062	0.051	0.044	0.048	0.057	0.061	0.059
Exp-F	0.041	0.071	0.094	0.058	0.047	0.067	0.047	0.054	0.06
C	0.028	0.05	0.094	0.036	0.044	0.082	0.034	0.037	0.073
C_{het}	0.043	0.066	0.094	0.051	0.058	0.096	0.046	0.044	0.083
H	0.155	0.163	0.189	0.126	0.119	0.156	0.097	0.086	0.123
H_{het}	0.078	0.082	0.099	0.082	0.07	0.067	0.066	0.049	0.057

From Table 2 it can be inferred that measurement error is has a very small influence to the performance of the tests considered if the given smooth change is either monotonic or non-monotonic. In small samples, hardly any of the tests shows any reaction to measurement error. But as the sample size increases, marginal power losses become apparent. The Hausman-type test seems to be affected most by an error margin in the explanatory variable. Still, the dimension of the power loss is at most four percentage points, a magnitude that can be considered meaningless. The testing procedures appear especially robust in the case of a non-monotonous smooth shift, as test powers may even increase slightly in some cases.

Table 2: Test powers with measurement error

	# obs	128			256			512		
	nts ratio	*0*	*0.15*	*0.3*	*0*	*0.15*	*0.3*	*0*	*0.15*	*0.3*
Change type	test									
monotonous	*Sup-F*	0.138	0.135	0.134	0.247	0.231	0.221	0.504	0.492	0.491
	Ave-F	0.111	0.111	0.116	0.171	0.164	0.168	0.369	0.364	0.361
	Exp-F	0.132	0.126	0.131	0.223	0.199	0.198	0.471	0.469	0.456
	C	0.075	0.071	0.07	0.148	0.142	0.151	0.366	0.358	0.35
	C_{het}	0.113	0.114	0.11	0.2	0.196	0.182	0.412	0.397	0.392
	H	0.298	0.287	0.281	0.404	0.391	0.378	0.634	0.61	0.606
	H_{het}	0.184	0.175	0.167	0.304	0.289	0.279	0.563	0.533	0.522
non-monotonous	*Sup-F*	0.151	0.159	0.156	0.309	0.313	0.314	0.681	0.671	0.658
	Ave-F	0.094	0.108	0.103	0.183	0.19	0.18	0.521	0.529	0.507
	Exp-F	0.143	0.136	0.144	0.272	0.297	0.281	0.657	0.663	0.657
	C	0.11	0.118	0.118	0.27	0.259	0.258	0.619	0.636	0.634
	C_{het}	0.153	0.166	0.164	0.34	0.321	0.315	0.678	0.684	0.695
	H	0.391	0.396	0.385	0.601	0.604	0.599	0.875	0.896	0.888
	H_{het}	0.252	0.264	0.259	0.483	0.502	0.492	0.822	0.848	0.833
high frequency	*Sup-F*	0.105	0.088	0.083	0.199	0.183	0.176	0.287	0.264	0.26
	Ave-F	0.068	0.072	0.066	0.121	0.117	0.116	0.167	0.156	0.149
	Exp-F	0.097	0.084	0.082	0.173	0.158	0.155	0.252	0.235	0.219
	C	0.106	0.094	0.079	0.163	0.171	0.151	0.26	0.251	0.223
	C_{het}	0.148	0.122	0.111	0.197	0.218	0.196	0.292	0.276	0.258
	H	0.282	0.254	0.225	0.371	0.357	0.344	0.462	0.437	0.402
	H_{het}	0.137	0.127	0.12	0.216	0.231	0.211	0.292	0.267	0.253

Looking at the scenario that involves a highly frequent smooth shift reveals a stronger relation between the extent of measurement error and test powers. The negative impact is strongest for the tests of Chen and Hong (2012a), although the evidence is mixed at medium sized samples. By contrast, the *ave-F* test experiences the smallest power losses. But generally and similarly to previous observations, the impact of measurement error on powers increases with the sample size.

Table 3: Powers in the presence of outliers

	# obs	128			256			512		
	outmag	*0*	*5*	*10*	*0*	*5*	*10*	*0*	*5*	*10*
Change type	test									
monotonous	*Sup-F*	0.138	0.108	0.128	0.247	0.163	0.118	0.504	0.392	0.232
	Ave-F	0.111	0.079	0.066	0.171	0.122	0.089	0.369	0.292	0.19
	Exp-F	0.132	0.102	0.105	0.223	0.138	0.108	0.471	0.36	0.225
	C	0.075	0.059	0.094	0.148	0.102	0.1	0.366	0.266	0.173
	C_{het}	0.113	0.081	0.097	0.2	0.127	0.12	0.412	0.311	0.197
	H	0.298	0.211	0.198	0.404	0.281	0.192	0.634	0.517	0.327
	H_{het}	0.184	0.119	0.103	0.304	0.183	0.103	0.563	0.425	0.252
non-monotonous	*Sup-F*	0.151	0.144	0.138	0.309	0.245	0.16	0.681	0.534	0.354
	Ave-F	0.094	0.081	0.066	0.183	0.131	0.081	0.521	0.39	0.208
	Exp-F	0.143	0.127	0.118	0.272	0.208	0.136	0.657	0.53	0.339
	C	0.11	0.076	0.094	0.27	0.159	0.109	0.619	0.49	0.265
	C_{het}	0.153	0.104	0.094	0.34	0.212	0.129	0.678	0.546	0.292
	H	0.391	0.279	0.204	0.601	0.44	0.238	0.875	0.78	0.52
	H_{het}	0.252	0.159	0.107	0.483	0.33	0.153	0.822	0.713	0.427
high frequency	*Sup-F*	0.105	0.087	0.108	0.199	0.122	0.088	0.287	0.208	0.138
	Ave-F	0.068	0.072	0.059	0.121	0.092	0.061	0.167	0.123	0.088
	Exp-F	0.097	0.084	0.093	0.173	0.104	0.086	0.252	0.178	0.126
	C	0.106	0.078	0.094	0.163	0.119	0.106	0.26	0.211	0.14
	C_{het}	0.148	0.093	0.092	0.197	0.156	0.119	0.292	0.24	0.168
	H	0.282	0.206	0.204	0.371	0.275	0.2	0.462	0.358	0.257
	H_{het}	0.137	0.111	0.104	0.216	0.162	0.096	0.292	0.226	0.151

When considering outliers instead of measurement error, a substantial negative effect on the tests power can be observed (Table 3). The pattern that can be observed depends largely on the sample size. In small samples, the presence of outliers itself has a negative impact on the powers of the test, whereas the magnitude of the outliers has only a small impact. In fact, the powers of the tests either converge towards some fixed value as the outliers become larger or increase with increasing outlier magnitude. By contrast, in large samples significant and strictly monotonous power losses can be observed for all tests as the outlier magnitude goes from 0 to 10. Most strikingly, the heteroskedasticity-robust Hausman-type tests statistic of Chen and Hong (2012a) has the largest power losses, despite the good behavior of its sizes. This power loss is probably due to the construction of the test statistic itself, as actually both Hausman-type test statistics experience similar power losses that exceed those of all other tests. It can also be observed that the power of each test increases with sample size at given outlier magnitude. One explanation is that the tests are less sensitive towards outliers at large sample sizes. Alternatively, the positive relation can simply be a result of a declining number of outliers relative to the entire sample size. In order to study this point further, the absolute number of outliers is changed from 4 to 8 and then 16 in a simulation study that is based on a sample of 512 observation and an outlier magnitude of ten standard deviations. As can be observed from Table 4, the results are mixed for test sizes. No clear tendency can be observed for the size of the concerned tests when the number of outliers changes. Hence it can be assumed that size is solely dependent on the magnitude of the outliers.

Table 4: Size and powers for an outlier magnitude of $z_2 = 10$

	# obs	512		
	# outliers	*4*	*8*	*16*
Change type				
none (sizes)	*Sup-F*	0.063	0.068	0.068
	Ave-F	0.059	0.064	0.048
	Exp-F	0.06	0.072	0.062
	C	0.073	0.056	0.085
	C_{het}	0.083	0.07	0.101
	H	0.123	0.11	0.147
	H_{het}	0.057	0.06	0.056
monotonous	*Sup-F*	0.232	0.151	0.086
	Ave-F	0.19	0.109	0.058
	Exp-F	0.225	0.14	0.077
	C	0.173	0.093	0.092
	C_{het}	0.197	0.109	0.11
	H	0.327	0.196	0.158
	H_{het}	0.252	0.123	0.082
non-monotonous	*Sup-F*	0.354	0.223	0.102
	Ave-F	0.208	0.115	0.056
	Exp-F	0.339	0.199	0.097
	C	0.265	0.144	0.094
	C_{het}	0.292	0.167	0.113
	H	0.52	0.282	0.181
	H_{het}	0.427	0.198	0.096
high frequency	*Sup-F*	0.138	0.081	0.068
	Ave-F	0.088	0.067	0.046
	Exp-F	0.126	0.085	0.063
	C	0.14	0.07	0.095
	C_{het}	0.168	0.091	0.11
	H	0.257	0.164	0.157
	H_{het}	0.151	0.083	0.073

A different impression is obtained from the simulated test powers. These decline unanimously when the relative number of outliers in the sample is raised. Only the Chow-based tests appear to be indifferent to the number of outliers when a monotonous or highly frequent shift is given and if more than eight outliers are given. In fact, the powers that are obtained with 16 outliers in the sample, a ratio that is equal to that of 4 outliers in a sample of 128 observations, are for the most part slightly below the comparator values that are given in the small sample. This implies that the test actually perform slightly worse in large samples than if the number of observations is small. The only exception to this observation are, again, the Chow-based tests of Chen and Hong (2012a), which perform similarly in samples of different sizes when the relative number of outliers is held constant or can even improve their performance slightly in large samples.

Still, the clear tendency of test powers to decrease in the presence of outliers is astonishing, as outliers may be expected to promote the detection of structural change due to the bias they introduce in coefficient estimators. But regardless of whether the observed development matches initial expectations or not, the severe power losses call for an adjustment procedure to eliminate outliers from the sample at hand.

4 Introducing wavelet-based outlier detection

In order to remove outliers from the data, an outlier-detection algorithm of Grané and Veiga (2010) is applied. The method is chosen for its emphasis on the residuals of an estimated econometric model and because it requires the model to be estimated only once, as opposed to e.g. Tsay (1986) who suggests an iterative procedure. Furthermore, the algorithm does not derive its threshold value from rules of thumb or asymptotical probabilistic reasoning (as e.g. Bilen and Huzurbazar, 2002), but obtains critical values for model residuals through simulation of similar processes of the same size. The simulation-based derivation of critical values allows as well to assume different distribution in case the standard assumption of normally distributed residuals is questionable. Finally, the algorithm addresses exclusively patterns of small range in the underlying data by employing the wavelet representation of a sequence. In this way, it restricts the scope to characteristics in the sample that are relevant for outliers and hence reduces the probability of errors.

Wavelet methods generally exploit a transformation that turns a time-dependent sequence into coefficients that capture activity at different frequency bands. This property is analogous to Fourier analysis, but unlike the infinitely oscillating sinoids used in spectral theory, wavelet analysis builds upon small wavelike functions that are nonzero in only a limited amount of time. The local scope of wavelets implies as well that, in contrast to Fourier analysis, information about the range *and* the position of a pattern in the sequence in question is captured by the resulting coefficients of the Discrete Wavelet Transform (DWT).

Due to two important results in Fourier analysis, only a small fraction of all possible frequencies has to be considered. Considering the Discrete Fourier Transform $\{\mathcal{X}(\frac{k}{N})\}$ of a finite time series $\{X_t\}$, where

$$\mathcal{X}(\frac{k}{N}) = \sum_{t=1}^{T} X_t e^{-i2\pi\frac{k}{N}(t-1)} \tag{4.1}$$

it generally holds that $\mathcal{X}(\frac{k}{N} + m) = \mathcal{X}(\frac{k}{N}) \ \forall \ m \in \mathbb{Z}$. Hence, it is only necessary to consider frequencies over a unit interval. A convenient choice is the Nyqvist frequency $\frac{k}{N} \in [-\frac{1}{2}; \frac{1}{2}]$, which has also logical appeal as the cyclic movements with the highest frequency that can be observed in a discrete sequence are patterns that spread over two adjacent observations. Moreover, it can be shown that $\mathcal{X}(\frac{k}{N}) = \mathcal{X}(-\frac{k}{N})$ as long as $\{X_t\} \in \mathbb{R}$. Consequently, the frequencies that have to be considered reduce to those in the interval $[0; \frac{1}{2}]$.

The DWT can be conducted in a number of ways of whom the fastest is the pyramid algorithm of Mallat (1989). The algorithm consists in applying a cascade of high-pass and low-pass filters to a time series that subsequently separate its characteristics at different frequency bands. A certain wavelet gives rise to a wavelet filter $\{h_l : l = 1, \ldots, L\}$ that satisfies the orthonormality conditions

$$\sum_{l=1}^{L} h_l = 0, \quad \sum_{l=1}^{L} h_l^2 = 1, \quad \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0 \, \forall \, n \in \mathbb{Z} \tag{4.2}$$

The wavelet filter is an approximation to a ideal high pass filter that extracts the upper half of the entire frequency band. A corresponding low-pass filter, the scaling filter, of the same length L can be derived from $\{h_l\}$ by applying the quadrature mirror relationship:

$$g_l = (-1)^l h_{L-l+1} \tag{4.3}$$

Applying the two filters to a certain sequence thus amounts to splitting its frequency band in half. Filtering is achieved by circularly convoluting the input sequence at hand with the corresponding filter and subsampling the result by two, i.e retaining only the odd-indiced elements of the filter output. This results in two $(\frac{T}{2} \times 1)$ vectors $\mathbf{W_1}$ and $\mathbf{V_1}$ that contain patterns at frequencies in the intervals $[\frac{1}{4}, \frac{1}{2}]$ and $[0, \frac{1}{4}]$ respectively and whose j-th element is given by

$$W_{1,j} = \sum_{l=1}^{T} h_l^{\circ} X_{2j-l \mod T}, \quad j = 1, \ldots, \frac{T}{2} \tag{4.4}$$

$$V_{1,j} = \sum_{l=1}^{T} g_l^{\circ} X_{2j-l \mod T}, \quad j = 1, \ldots, \frac{T}{2} \tag{4.5}$$

where $\{h_l^\circ\}$ and $\{g_l^\circ\}$ are the wavelet and scaling filters periodised to the length T of the sequence and mod T implies that the series $\{X_t : t = 1, \ldots, T\}$ is periodically repeating outside of its borders, i.e. $X_{mT+t} = X_t \; \forall \; m \in \mathbb{Z}, t = 1, \ldots, T$. Given that a pattern with frequency $\frac{1}{2}$ spans over 2 observations, each DWT coefficient $W_{1,j}$ captures the variation over the two observations X_{2j} and X_{2j-1}. The pyramid algorithm of Mallat (1989) conducts the filtering for the initial sequence and subsequently for the obtained scaling coefficients. In this way, at k-th stage it summarizes the contribution of the time series over the frequency band $[1/2^{k+1}; 1/2^k]$ in a column vector $\mathbf{W_k}$ that contains $T/2^k$ coefficients. Obviously, the pyramid algorithm hence requires that the sample size is dyadic[1]: If the time series consists of $T = 2^K$ observations, the pyramid algorithm has K stages. This also implies that the low pass filter output at stage K, summarizing the frequency band $[0; 1/2^{K+1}]$, is not processed further.

The DWT coefficients that result from applying the pyramid algorithm are kept in a $(T \times 1)$-vector $\mathbf{W}$ that stacks the wavelet coefficient vectors from all stages of the pyramid algorithm plus the scaling coefficient from its last stage. Summarizing the transform operations in a matrix $\mathcal{W}$, the DWT coefficients can be obtained by matrix multiplication:

$$\mathbf{W} = \mathcal{W}\mathbf{X} \tag{4.6}$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{W_1} \\ \vdots \\ \mathbf{W_K} \\ \mathbf{V_K} \end{bmatrix} \tag{4.7}$$

Furthermore, given the orthonormality conditions of the wavelet filters the original time series can be reconstructed from the DWT coefficients by pre-multiplying the latter with $\mathcal{W}^T$, as $\mathcal{W}^T\mathbf{W} = \mathcal{W}^T\mathcal{W}\mathbf{X} = \mathbf{X}$. Another implication of the orthonormality conditions is that the DWT coefficients preserve the energy of the initial sequence:

$$\|\mathbf{X}\|^2 = \sum_{t=1}^{T} X_t^2 = \sum_{k=1}^{K} \sum_{j=1}^{\frac{T}{2^k}} W_{k,j}^2 = \|\mathbf{W}\|^2 \tag{4.8}$$

For a more extensive treatment of the DWT using the same notation, the interested reader is referred to Percival and Walden (2002).

Analyzing DWT coefficients is generally extremely valuable when the feature of interest within a time series is characterized by activity at a certain limited frequency band. Information in the time domain, as given by an initial untransformed series, cannot capture these features separately from irrelevant patterns at other frequencies. Without making a frequency-based distinction, it is possible to be taken in by spurious patterns or to fail to detect interesting characteristics that are masked by variation at other frequency bands. The possibility to select the information of interest from DWT coefficients at different scales hence constitutes a clear advantage of wavelet-based methods over procedures that purely rely on information from the time domain.

Outliers in the time-dimension of a given sequence can be seen as variation at very high frequency, as they span only over one single observation. Consequently, they will have an impact on the finest scale coefficients of the wavelet transform . Due to the energy preserving property of the DWT, the wavelet coefficients that capture an outlier in the initial series will be characterized by a high absolute value. In addition, as the DWT provides information about

[1] This restriction can be weakened by conducting the Partial Discrete Wavelet Transform which manually imposes number of stages $\bar{k} < K$

variation at different frequency bands and the position of this variation in the time domain, it is possible to localize an outlier from the DWT coefficients.

Outlier detection according to the method of Grané and Veiga (2010) is based upon the finest-scale DWT coefficients of the residuals $\{e_t : t = 1, \ldots, T\}$ of a chosen econometric model and hence addresses precisely the set of information that is assumed to contain information about the position of outliers. The distributional properties of model residuals provide generally an excellent basis for outlier detection, as it is solely required to investigate which residuals do not fit into an i.i.d distribution around a mean of zero. The algorithm employs a threshold rule to distinguish outliers from ordinary errors, where the threshold itself is the 95% percentile for the maximum of simulated normally distributed series of length T. The threshold choice is motivated by the assumption that the residuals of a correctly specified model should be white noise and the fact that the DWT as an orthonormal transform preserves the distributional properties of the underlying time series (Percival and Walden, 2002). All coefficients whose absolute value exceeds the threshold are considered outliers and their indice is noted, which results in a set $S = \{s_1, \ldots, s_N\}$ of identified outlier positions. Outlier correction is conducted in the time domain. As each finest-scale DWT coefficient spans over two observations of the initial sequence, the corresponding time-domain values are compared to the average of all other elements of the sequence $\bar{e}_{T-2}$. The observation that deviates most from the average is identifier as the outlier and set to zero. This procedure is summarized by the following decision rule:

$$e_{2s_n} = \begin{cases} e_{2s_n} & \text{if } |e_{2s_n} - \bar{e}_{T-2}| < |e_{2s_n-1} - \bar{e}_{T-2}| \\ 0 & \text{if } |e_{2s_n} - \bar{e}_{T-2}| \geq |e_{2s_n-1} - \bar{e}_{T-2}| \end{cases} \tag{4.9}$$

and

$$e_{2s-1_n} = \begin{cases} e_{2s-1_n} & \text{if } |e_{2s_n-1} - \bar{e}_{T-2}| \leq |e_{2s_n} - \bar{e}_{T-2}| \\ 0 & \text{if } |e_{2s_n-1} - \bar{e}_{T-2}| > |e_{2s_n} - \bar{e}_{T-2}| \end{cases} \tag{4.10}$$

where $s_n \in S$ is the position of a DWT coefficient that exceeded the threshold.

It is desirable to implement the algorithm into the simulation study in order to assess whether it helps compensating the power losses that are due to outliers without impairing the results too much in cases where outliers are not present. In the simulation study, wavelet based outlier detection is conducted using the Haar wavelet, the simplest possible wavelet that gives rise to the wavelet and scaling filters

$$h_l = \left\{ -\frac{1}{\sqrt{2}}; \frac{1}{\sqrt{2}} \right\} \text{ and } g_l = \left\{ \frac{1}{\sqrt{2}}; \frac{1}{\sqrt{2}} \right\} \tag{4.11}$$

As depticted in Table 5, applying the algorithm to data that do not have any outliers results only in very marginal size decreases. Hence, the outlier elimination technique is sensitive enough to maintain the underlying data. When the magnitude of the outliers is increased subsequently, the test sizes of the corrected data do not increase substantially, as it was the case for the unadjusted samples. In most cases, only minuscule changes upwards can be observed, but these can be considered negligible. Only the homeskedasticity-specific version of the Hausman test experiences size losses that are significant at the lowest sample size. But given the high initial size of the test, such an adjustment is not alarming.

Table 5: Sizes of corrected data

# obs	128				256				512			
outmag	0*	0	5	10	0*	0	5	10	0*	0	5	10
test												
Sup-F	0.043	0.04	0.065	0.053	0.058	0.054	0.057	0.054	0.051	0.055	0.059	0.059
Ave-F	0.036	0.046	0.059	0.048	0.051	0.049	0.049	0.05	0.057	0.047	0.053	0.052
Exp-F	0.041	0.039	0.061	0.051	0.058	0.051	0.056	0.052	0.047	0.057	0.065	0.063
C	0.028	0.026	0.049	0.043	0.036	0.025	0.039	0.032	0.034	0.025	0.034	0.028
C_{het}	0.043	0.045	0.067	0.056	0.051	0.044	0.052	0.049	0.046	0.036	0.041	0.037
H	0.155	0.139	0.174	0.151	0.126	0.104	0.123	0.111	0.097	0.077	0.083	0.075
H_{het}	0.078	0.07	0.089	0.079	0.082	0.065	0.073	0.076	0.066	0.05	0.059	0.052

$*$ = uncorrected data

With regards to the powers of the tests for smooth structural change, the algorithm of Grané and Veiga (2010) leads only to small power losses in samples where outliers are not given. The exact results are quite mixed, but generally the power losses do not exceed 5 percentage points. Notably the Ave-F test and both Hausman-type tests appear to be robust to the data correction in small and medium size samples, as the power declines only marginally or increases slightly. The powers for samples that include outliers with a magnitude of 5 or 10 times the standard deviation of the dependent variable are restored almost entirely by the data correction mechanism, irrespectively of which test is considered. For an outlier magnitude of five standard deviations, power losses in some medium size and most large samples exist with respect to the comparator values. This indicates that Grané and Veiga's (2010) algorithm fails to detect a small number of outliers in these cases. However, test powers for larger outliers generally approach those of the comparator samples, although their values still may lie up to 5 percentage points below in the scenario of a single monotonous gradual change.

Table 6: Test powers of adjusted samples

	# obs	128				256				512			
	outmag	0*	0	5	10	0*	0	5	10	0*	0	5	10
Change type	test												
monotonous	*Sup-F*	0.138	0.104	0.136	0.119	0.247	0.224	0.219	0.223	0.504	0.469	0.432	0.47
	Ave-F	0.111	0.104	0.12	0.112	0.171	0.173	0.167	0.164	0.369	0.339	0.322	0.341
	Exp-F	0.132	0.111	0.138	0.121	0.223	0.199	0.201	0.205	0.471	0.431	0.404	0.426
	C	0.075	0.068	0.091	0.079	0.148	0.142	0.147	0.147	0.366	0.336	0.312	0.319
	C_{het}	0.113	0.095	0.117	0.104	0.2	0.189	0.192	0.189	0.412	0.392	0.372	0.38
	H	0.298	0.285	0.288	0.287	0.404	0.384	0.367	0.382	0.634	0.587	0.566	0.593
	H_{het}	0.184	0.175	0.185	0.166	0.304	0.275	0.266	0.269	0.563	0.513	0.477	0.513
non-monotonous	*Sup-F*	0.151	0.141	0.19	0.155	0.309	0.323	0.299	0.312	0.681	0.647	0.607	0.634
	Ave-F	0.094	0.105	0.122	0.111	0.183	0.198	0.173	0.186	0.521	0.488	0.433	0.478
	Exp-F	0.143	0.144	0.172	0.149	0.272	0.297	0.277	0.291	0.657	0.62	0.588	0.624
	C	0.11	0.112	0.129	0.119	0.27	0.255	0.242	0.245	0.619	0.597	0.569	0.602
	C_{het}	0.153	0.158	0.177	0.166	0.34	0.33	0.31	0.312	0.678	0.655	0.625	0.645
	H	0.391	0.401	0.411	0.385	0.601	0.615	0.571	0.605	0.875	0.873	0.835	0.862
	H_{het}	0.252	0.268	0.274	0.257	0.483	0.492	0.453	0.474	0.822	0.82	0.772	0.811
high frequency	*Sup-F*	0.105	0.112	0.117	0.106	0.199	0.174	0.169	0.176	0.287	0.257	0.258	0.258
	Ave-F	0.068	0.073	0.09	0.086	0.121	0.107	0.106	0.105	0.167	0.139	0.137	0.141
	Exp-F	0.097	0.106	0.114	0.111	0.173	0.153	0.151	0.153	0.252	0.213	0.223	0.233
	C	0.106	0.099	0.112	0.1	0.163	0.161	0.172	0.173	0.26	0.247	0.247	0.245
	C_{het}	0.148	0.126	0.137	0.126	0.197	0.199	0.209	0.206	0.292	0.276	0.276	0.268
	H	0.282	0.257	0.275	0.261	0.371	0.361	0.356	0.358	0.462	0.426	0.423	0.423
	H_{het}	0.137	0.131	0.15	0.133	0.216	0.219	0.22	0.223	0.292	0.257	0.258	0.253

$*$ = uncorrected data

5 Conclusion

This thesis has studied the effect of two types of data inaccuracies on the size and power of test sfor smooth structural change. It has been shown that random measurement error of a magnitude that can be found in real economic data does not impair the performance of the considered tests. This results holds irrespectively of the sample size and the type of structural change. As the chosen error magnitudes can be seen as an upper bound to possible error magnitudes in available data, it can be said that tests for smooth structural change are robust to measurement error. In the unlikely case of substantially larger measurement error, however, similar detrimental effects of error margins in the explanatory variable, as observed by Li (2012), are likely to appear.

By contrast, outliers do have a significant impact by inflating the size and lowering the powers of tests for smooth structural change. The extent of these performance deterioration calls for an adjustment mechanism in order to remove outliers from the data. The algorithm of Grané and Veiga (2010) is conceptually suited to perform this adjustment as it focuses strictly on the unexplained portion of a chosen econometric model and uses wavelet techniques to address strictly high-frequency variations in the underlying series. Simulation results show that the algorithm indeed succeeds in compensating the power losses due to outliers to a large extent, although its performance is not perfect in large samples and in cases where the outliers do not differ excessively from regular observations.

References

D. W. K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–56, July 1993.

D. W. K. Andrews and W. Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6):1383–1414, November 1994.

N. S. Balke and T. B. Fomby. Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series. *Journal of Applied Econometrics*, 9(2):pp. 181–200, 1994.

P. Bauer and P. Hackl. The use of mosums for quality control. *Technometrics*, 20(4):pp. 431–436, 1978.

C. Bilen and S. Huzurbazar. Wavelet-based detection of outliers in time series. *Journal of Computational and Graphical Statistics*, 11(2):pp. 311–327, 2002.

J. Bound, C. Brown, and N. Mathiowetz. Measurement error in survey data. In J. Heckman and E. Leamer, editors, *Handbook of Econometrics*, volume 5 of *Handbook of Econometrics*, chapter 59, pages 3705–3843. Elsevier, January 2001.

R. L. Brown, J. Durbin, and J. M. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2): pp. 149–192, 1975.

B. Chen and Y. Hong. Testing for smooth structural changes in time series models via nonparametric regression. Working Paper, University of Rochester, 2008.

B. Chen and Y. Hong. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183, May 2012a.

B. Chen and Y. Hong. Supplemental material for *Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression*. `http://www.econometricsociety.org/suppmat.asp?id=383&vid=80&iid=3&aid=8`, May 2012b. [Online; accessed 05-Aug-2012].

G. C. Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):pp. 591–605, 1960.

S. Dees, F. d. Mauro, M. H. Pesaran, and L. V. Smith. Exploring the international linkages of the euro area: a global var analysis. *Journal of Applied Econometrics*, 22(1):1–38, 2007. ISSN 1099-1255.

J. U. Farley, M. Hinich, and T. W. McGuire. Some comparisons of tests for a shift in the slopes of a multivariate linear time series model. *Journal of Econometrics*, 3(3):297 – 318, 1975. ISSN 0304-4076.

A. J. Fox. Outliers in time series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(3):pp. 350–363, 1972.

A. Grané and H. Veiga. Wavelet-based detection of outliers in financial time series. *Computational Statistics & Data Analysis*, 54(11):2580 – 2593, 2010.

B. E. Hansen. The new econometrics of structural change: dating breaks in u.s. labour productivity. *Journal of Economic Perspectives*, 15(4):117–128, Fall 2001.

H.-J. Kim and D. Siegmund. The likelihood ratio test for a change-point in simple linear regression. *Biometrika*, 76(3):pp. 409–423, 1989.

T. Laubach. New evidence on the interest rate effects of budget deficits and debt. *Journal of the European Economic Association*, 7(4):858–885, 2009. ISSN 1542-4774.

Y. Li. Investigation of the structural change test with measurement error - a bayesian wavelet based improvement. presentation at departmental seminar, Lunds Universitet, Department of Economics, 27th April 2012.

C.-F. J. Lin and T. Teräsvirta. Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, 62(2):211–228, June 1994.

R. E. Lucas. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19 – 46, 1976. ISSN 0167-2231.

S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

M. M. McConnell and G. Perez-Quiros. Output fluctuations in the united states: What has changed since the early 1980's? *The American Economic Review*, 90(5):pp. 1464–1476, 2000. ISSN 00028282.

O. Morgenstern. *On the accuracy of economic observations*. Princeton University Press, Princeton, New Jersey, 1963.

D. Percival and A. Walden. *Wavelet methods for time series analysis*. Cambridge Series In Statistical And Probabilistic Mathematics. Cambridge University Press, 2002. ISBN 9780521685085.

R. E. Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53(284):pp. 873–880, 1958.

R. E. Quandt. Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, 55(290):pp. 324–330, 1960. ISSN 01621459.

J. Stock and M. Watson. *Introduction to econometrics*. Addison-Wesley Series in Economics. Pearson/Addison Wesley, 2007. ISBN 9780321442536.

R. S. Tsay. Time series model specification in the presence of outliers. *Journal of the American Statistical Association*, 81(393):pp. 132–141, 1986.

A. Zeileis, F. Leisch, K. Hornik, and C. Kleiber. `strucchange`: an R package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7(2):1–38, 2002. ISSN 1548-7660.