Test of Equality Between Regression Lines in Presence of Errors in Variables

Zhang Fan

Supervisor: Bo Wallentin

Department of Statistics, Uppsala University

June 4, 2012
Abstract

Testing of equality between regression lines is a common procedure when people want to see if there is a structural change in econometric time series. In this paper, we want to discuss the reliability of this procedure in presence of error in variables. In our study, a single explanatory variable model, vector explanatory variables model, instrumental variable model and first difference model are provided. We simulate empirical sizes for the Chow Test, the predictive Chow test and the CUSUM Test. The results show that the empirical sizes are equal to the magnitudes of the significance levels for first difference model with trending regressors. However in other situations, the sizes are far too high from the stated levels for the Chow test and the predictive Chow test. The proportions of rejections for the CUSUM test often preforms too low, implying a loss of power for this test.

*Keywords:* Error-in-variables; Regression model; Test of stability; First difference model
1 Introduction

Testing the equality between regression lines is a common procedure in applied econometrics. This could be used when we want to test the equality of the harvest of the crops in different areas or the stability of the harvest in time series. More literatures can be found in Hackl and Westlund (1989) which give an introduction about the structural change. Most time series and economic data have measurement errors, but these errors are often ignored in practice. One reason is that there is little progress for the effects on testing in presence of error in variables. Also the lack of information of the errors prevent the economist from taking measurement errors into consideration. Eklöf (1992) has reported the test for structural change where part of the data are "preliminary". After some simulation he reports that the empirical sizes for the Chow test (Chow, 1960) are too large compared with the significance levels. Jonsson (1997) has tested for structural change with the Chow test and the CUSUM test when there are measurement errors in both regressors and dependent variables. He chose different variances for the errors and make a simulation study for a simple model with two kind of regressors: normally distributed regressor and trending regressor. The results show that for the Chow test the empirical sizes are too high especially when one of the subsets is relative small. There are many reasons why the Chow test which is widely used for tests of stability has a bad performance in presence of errors in variables. Measurement errors in regressors could causes ordinary least squares (OLS) estimators to be biased and inconsistent because these estimators are affected by the errors in regressors and errors in equation. We could have problem in testing the stability since the bias of the estimators will cause the size to be too high for the Chow test. Also the measurement error in variables could lead to heteroscedasticity which could cause the size to be too large for the Chow test. See Toyoda (1974) and Schmidt and Sickles (1977) for more details.

In this paper we will discuss how do measurement errors affect the structural change test and show what would happen when we meet errors-in-variables. We will consider four type of models: one regressor model, two regressors model, first difference model and instrumental variable model. Also we will use three types of test: the Chow test, the predictive Chow test and the CUSUM test.
2 Measurement error in regression line

In this section we will introduce the linear model with measurement errors and see the effect of measurement error for OLS estimators. A classical linear regression model with one independent variables is shown by:

\[ Y_t = \beta_1 + \beta_2 x_t + u_t, \quad t = 1, 2, \ldots, T \quad (1) \]

where errors \( u_t \) are independently normally distributed with mean 0 and variance \( \sigma_u^2 \). Here \( x_t \) is considered to be fixed in repeated sampling. Based on assumption that \( \text{Cov}(x, u) = 0 \), then the Ordinary Least Squares estimator for the slope \( \beta_2 \) will be:

\[ \hat{\beta}_2 = \left[ \sum_{t=1}^{n} (x_t - \bar{x})^2 \right]^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(Y_t - \bar{Y}) \]

Now we suppose that we can not observe the true value of \( x_t \) directly, instead we can observe \( X_t \), which can be explained as the true \( x_t \) plus some measurement error. That is:

\[ X_t = x_t + \varepsilon_t \quad (2) \]

where \( \varepsilon_t \) is a random variable with mean 0 and variance \( \sigma^2_\varepsilon \) and \( \varepsilon_t \) is uncorrelated with \( x_t \). The observed variable \( X_t \) can be called indicator variable while the unobserved variable \( x_t \) is called latent variable. The model (1) is called structure equation and (2) is called measurement error equation.

As an example of this situation that the true value of \( x_t \) can’t be observed, one can consider the relationship between output of marsh gas and the number of microorganism which produce the marsh gas in a certain marsh area. Suppose (1) is an approximation of the relationship between this gas and microbe. \( \beta_2 \) is the expected increase in volume of marsh gas when the microbe goes up for one unit. To estimate the number of microorganism, it is necessary to get some samples from this area and do laboratory analysis on this samples. When doing this analysis, we can not observe true \( x_t \), but we can observe the estimate of \( x_t \). Here the observed \( X_t \) are equal to the true \( x_t \) plus measurement error which come from random sampling and laboratory analysis.

Now we concentrate on the effect of measurement errors when estimating the slope with ordinary least square(OLS). Consider the regressor \( x_t \) to be a random variable, assume that \( x_t, u_t \) and \( \varepsilon_t \) are independent and each of them follows a normal distributions with variances \( \sigma_x^2, \sigma_u^2 \) and \( \sigma_\varepsilon^2 \), respectively.
The least squares regression coefficient for the slope based on observed $X_t$ is known to be

$$
\hat{\gamma}_2 = \left[ \sum_{t=1}^{n} (X_t - \bar{X})^2 \right]^{-1} \sum_{t=1}^{n} (X_t - \bar{X})(Y_t - \bar{Y}) \quad (3)
$$

The expected value of the estimator then can be shown to be (see Fuller 1987):

$$
E\hat{\gamma}_2 = \beta_2(\sigma_x^2 + \sigma^2_\varepsilon)^{-1}\sigma_x^2 \quad (4)
$$

The ratio $K_{xx} = \sigma_x^2(\sigma_x^2 + \sigma^2_\varepsilon)^{-1}$ is called reliability ratio and it can describe the effect of measurement error on the estimations of $\beta_2$.

Now consider the model with multiple regressors:

$$
Y = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + u_t \quad t = 1, 2, \ldots, T \quad (5)
$$

where $y, x_1, x_2, \ldots, x_{k-1}$ are observable but $x_k$ is not. Instead we observe $X_k$ which is a measure of $x_k$. The measurement error model is:

$$
X_{kt} = x_{kt} + \varepsilon_t \quad t = 1, 2, \ldots, T \quad (6)
$$

where $x_k$ and $u_t$ are uncorrelated and $\varepsilon_t$ is a random variable with mean 0 and variance $\sigma_\varepsilon$. Based on the general classical errors in variables (CEV) assumption $\text{Cov}(x_k, \varepsilon_k) = 0$, the probability limit of OLS estimators is as follows. Details can be found in Levi (1973):

$$
\text{plim} \hat{\beta}_{OLS} = (\Sigma + \Omega)^{-1}\Sigma \beta \quad (7)
$$

where $\Sigma$ and $\Omega$ are the variance-covariance matrices for true regressors and measurement errors and $\hat{\beta}_{OLS}$ is the vectors of OLS estimators for the parameters of the regressors. When only $x_k$ is measured with errors, the probability limits for the OLS estimator for regressors $\beta$ are:

$$
\text{plim} \hat{\beta}_k = \frac{\sigma_{r_k}^2}{\sigma_{e_k}^2 + \sigma_{r_k}^2}\beta_k, \text{ and } \text{plim} \hat{\beta}_i = \beta_i + \frac{\sigma_{r_k}^2}{\sigma_{e_k}^2 + \sigma_{r_k}^2}\gamma_i \beta_k, \ i \neq k \quad (8)
$$

where $r_k$ is the error for the regression of true $x_k$ on the remaining regressors $x_i, i \neq k$, and the unobserved $x_k$ should be uncorrelated with the observed $x_i, i \neq k$. $\gamma_i$ is the parameter of $x_i$ in the regression of $x_k$. More details can be found in Wooldridge (2010:76-82). From the equation we can see that it is the variance in $x_k$ after linked with other regressors that affects the probability limits for $\hat{\beta}_k$. So if the unobserved $x_k$ is collinear with observed variables more, the $\hat{\beta}_k$ will be biased more.
In this paper we build four types of models which are applicable for econometric time series data: a simple model, a general model, a first difference model and an instrumental variable model. Each model have measurement errors. First two models are simplest forms of linear regression models. First difference model is used when there are strong autocorrelations for the error term. The instrumental variable model is used to get a consistent unbiased estimator for the regressors and see how do these estimators affect the structural change tests.

2.1 Model 1: Simple regression model

Assume the following model:

Structural equation:

\[ y_t = \beta_1 + \beta_2 x_t + q_t, \text{ where } q_t \text{ is } IN(0, \sigma^2_q), t = 1, \ldots, T \]  \hspace{1cm} (9)

Measurement equation:

\[ Y_t = y_t + w_t \quad X_t = x_t + \varepsilon_t \text{ where } w_t \text{ is } IN(0, \sigma^2_w) \text{ and } \varepsilon_t \text{ is } IN(0, \sigma^2_\varepsilon) \]

Also we assume that the errors \( w_t, q_t \) and \( \varepsilon_t \) are independent. \( \sigma^2_w \) stands for the variance of \( q_t \) plus \( w_t \). For the stochastic regressors, we assume that the mean and variance of the regressor are constant and they can be labeled \( \mu_x \) and \( \sigma^2_x \).

Let us see how the measurement errors in the regressor affect the test. Concentrating on slope estimator \( \hat{\beta}_2 \) we have:

\[
\text{plim} \hat{\beta}_2 = \beta_2 K, \text{ where } \hat{\beta}_2 = \frac{S_{xy}}{S_{xx}} = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2} \hspace{1cm} (10)
\]

where

\[ K = \frac{\sigma^2_x}{\sigma^2_x + \sigma^2_\varepsilon} \hspace{1cm} (11) \]

From (10) and (11) we can see that the asymptotic bias of \( \beta_2 \) in (9) is related to the reliability ratio \( K \). When there is no measurement error, the \( K \) goes up to 1. For further readings about this model, see Fuller(1987) for more details.

2.2 Model 2: General regression model

Assume the following model:
Structural equation:
\[ y_t = \beta_1 + \beta_2 x_{1t} + \beta_3 x_{2t} + q_t, \text{ where } q_t \text{ is } \mathcal{N}(0, \sigma_q^2), t = 1, \ldots T \] (12)

Measurement equation:
\[ Y_t = y_t + w_t \quad X_{2t} = x_{2t} + \varepsilon_t \text{ where } w_t \text{ is } \mathcal{N}(0, \sigma_w^2) \text{ and } \varepsilon_t \text{ is } \mathcal{N}(0, \sigma_\varepsilon^2) \]

We assume that the error \( w_t, q_t \) and \( \varepsilon_t \) are independent and \( x_1 \) and \( x_2 \) are uncorrelated. \( \sigma_q^2 \) stands for the variance of \( q_t + w_t \). The regressor need to have constant means and variances if they are stochastic regressors. Compared with the simple model, in this model we add a regressor which has no measurement errors. This maybe a more general case that we would meet in applied econometrics. In most way of dealing with structural break we ignore this measurement error, but in my study this measurement error can not be disregarded. From (8) we can see that in this model the OLS estimator for regressor will be:
\[
\text{plim} \hat{\beta}_2 = \beta_2 \text{ and } \text{plim} \hat{\beta}_3 = \frac{\sigma_{x_2}^2}{\sigma_{x_2}^2 + \sigma_\varepsilon^2} \beta_3
\] (13)

where \( \sigma_{x_2}^2 \) is the variance of the regressor \( x_2 \), \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are the OLS estimators for parameters of the regressors. Here the reliability ratio for \( x_2 \) can be shown as:
\[
K = \frac{\sigma_{x_2}^2}{\sigma_{x_2}^2 + \sigma_\varepsilon^2}
\] (14)

We can see that the reliability ratio for \( \hat{\beta}_3 \) perform the same result as it shows in model 1 and \( \beta_2 \) is consistently unbiased estimated. When the measurement errors goes up, the estimator will be far from the true value, making it difficult to get an accurate regression.

### 2.3 Model 3: First difference model

Assume the following model: \( y_t = \beta_1 + \beta_2 x_t + q_t \), where \( q_t \) are fully positive autocorrelated. Then we can take the first difference and get a first difference model:

Structural equation\(^1\):
\[
\nabla y_t = \beta \nabla x_t + p_t, \text{ where } p_t = q_t - q_{t-1}, \text{ and } p_t \text{ is } \mathcal{N}(0, \sigma_p^2), t = 1, \ldots T
\] (15)

\(^1\nabla\) is the first difference operator and \( \nabla y_t = y_t - y_{t-1} \).
Measurement equation:

\[ X_t = x_t + \varepsilon_t \text{ where } \varepsilon_t \text{ is } IN(0, \sigma^2_\varepsilon) \]

Note that the errors \( q_t \) are fully correlated. Here we assume that \( p_t \) and \( \varepsilon_t \) are independent. In this model we do not invite measurement errors in \( Y_t \). Because if measurement error in \( Y_t \) does exist, which is labeled \( w_t \), then after taking first difference the variances of the total error (given by \( \nabla u_t = \nabla w_t + p_t \)) would be correlated and the OLS assumptions would not be fulfilled. In this model the mean and variance of the regressor should be constant when the regressor is stochastic. Now this model can be regarded as model 1 with no intercept. So the asymptotic bias of \( \beta \) can be shown to be:

\[
\text{plim} \hat{\beta} = \beta K \quad \text{where } K = \frac{\text{Var}(x_t - x_{t-1})}{\text{Var}(x_t - x_{t-1}) + 2\sigma^2_\varepsilon} 
\]

(16)

Where \( K \) is the reliability ratio. The advantage of the first difference approach is that it removes the latent heterogeneity from the model. When the error variances have strong autocorrelation, after taking first difference, the model (15) has random normal disturbances, so we can use the Chow test, the predictive Chow test and the CUSUM test for first difference model. Take model 1 as a simple example. First we invite model 1 with fully autocorrelation in error variance: \( u_t = u_{t-1} + \eta_t \), where \( \eta_t \) follows a independent normal distribution. So when we take first difference, the model falls to model 3 and the error variance are i.i.d, then the test can be used. Also the first difference model is appropriate for trending regressor situation. Suppose in this model \( x_t \) is random walk with a drift: \( x_t - x_{t-1} = \mu + \rho_t \) where \( \mu \) is a constant and \( \rho_t \) is i.i.d normally distributed. Then after taking first difference the new regressor \( \nabla x_t \) will be normally distributed which makes it easier to do some inference.

2.4 Model 4: Instrumental variable model

Assume the model:

Structural equation:

\[ y_t = \beta_1 + \beta_2 x_t + u_t, \text{ where } u_t \text{ is } IN(0, \sigma^2_u), t = 1, \ldots, T \]  

(17)

Measurement equation:

\[ X_t = x_t + \varepsilon_t \text{ where } \varepsilon_t \text{ is } IN(0, \sigma^2_\varepsilon) \]
The errors $u_t$ and $\varepsilon_t$ are independent. For the stochastic regressors, we assume that the mean and variance of the regressor are constant and they can be labeled $\mu_x$ and $\sigma^2_x$. Here we have a variable $z_t$ that satisfy two conditions:

$$E \left\{ T^{-1} \sum_{t=1}^{T} (z_t - \bar{z})(u_t, \varepsilon_t) \right\} = (0, 0)$$

$$E \left\{ T^{-1} \sum_{t=1}^{T} (z_t - \bar{z})x_t \right\} \neq 0$$

where $\bar{z} = T^{-1} \sum_{t=1}^{T} z_t$. Then $z_t$ is called an instrumental variable for $x_t$ of this model. Then $x_t$ can be written as:

$$x_t = \delta_0 + \delta_1 z_t + \gamma_t, \quad t = 1, 2, \ldots, T$$

It follows that we can have unbiased consistent estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. $\hat{\beta}_2 = m_{XZ}^{-1} m_{YZ}$ and $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$

where $\bar{Y} = T^{-1} \sum_{t=1}^{T} Y_t$, $\bar{X} = T^{-1} \sum_{t=1}^{T} X_t$ and $$(m_{XZ}, m_{YZ}) = (T - 1)^{-1} \sum_{t=1}^{T} (Y_t - \bar{Y}, X_t - \bar{X})(z_t - \bar{z})$$

More details and a survey can be found by Fuller (1987).

3 Test of stability

The purpose of testing the stability is to see if there is a structural break or if the model is the same for different subsets of the data. The null hypothesis of interest is that the parameters of regressors in the models stay same for all the observations. The well known Chow test is a test of whether the coefficients in two linear regressions on different data sets are equal. First we estimate the regression model with all the observations and get the residual sum of squares ($RSS_R$), then we split the data into two parts. For each part, we fit the regression and obtain the sum of squares residuals ($RSS_1$ and $RSS_2$). Then the Chow test statistics is:

$$\frac{(RSS_R - (RSS_1 + RSS_2))/k}{(RSS_1 + RSS_2)/(N_1 + N_2 - 2k)} \quad (18)$$

where $k$ is the number of parameters, $N_1$ and $N_2$ are the number of observations in each group. If the null hypothesis is true, the test statistics follows
an F distribution with \( k \) and \( N_1 + N_2 - 2k \) degrees of freedom. The test requires that the error variances are the same for all observations and are independently normally distributed. If the variances are different, then the classical regression model no longer apply. Another requirement for the Chow test is that there are enough data for each group. In some circumstances, however, the data series are not long enough for one or the other group to fit a regression. In such a situation another test can be used to test the similar hypothesis:

\[
H_0: \quad E(y|\mathbf{X}; \beta_1) = E(y|\mathbf{X}; \beta_2) \quad \text{(19)}
\]

Then the test statistics is:

\[
\frac{(RSS_R - RSS_1)/N_2}{RSS_1/(N_1 - k)} \quad \text{(20)}
\]

Note that the degrees of freedom in numerator is \( N_2 \). This test is named the **predictive Chow test**. For linear model, the Chow test and Chow the predictive Chow test need a known single break in mean. Details of these two tests can be found in Greene(2012) If we do not provide the structural break point, a type of test the CUSUM test can be used for this situation.

The Brown-Durbin-Evans(1975) **CUSUM test** is used to test for a structural change in time series data. The advantage of this test is that no structural break point is needed. Take model 1 as an example. Suppose we have \( T \) observations. The \( r \)-th recursive residual for \( y_t \), also named "one step ahead prediction error" is:

\[
q_r = Y_r - \hat{\beta}_{1,r-1} - \hat{\beta}_{2,r-1}X_r \quad \text{(21)}
\]

where \( \hat{\beta}_{1,r-1} \) and \( \hat{\beta}_{2,r-1} \) are OLS estimator computed with first \( r - 1 \) observations. Then we can calculate the \( r \)-th scaled residual:

\[
w_r = \frac{q_r}{[1 + X_r'(X_{r-1}'X_{r-1})^{-1}X_r]^{1/2}} \quad \text{(22)}
\]

Under the null hypothesis that there is no structural changes, we can see that the scaled residuals follows a normal distribution with mean 0 and variance \( \sigma^2 \). So the CUSUM test can be provided with cumulated sum of residual:

\[
W_t = \sum_{r=K+1}^T \frac{w_r}{\hat{\sigma}_e} \quad \text{(23)}
\]

where

\[
\hat{\sigma}_e^2 = \frac{1}{T - (k + 1)} \sum_{r=k+1}^T (w_r - \bar{w})^2 \quad \text{and} \quad \bar{w} = \frac{1}{T - k} \sum_{r=k+1}^T w_r \quad \text{(24)}
\]

8
The null hypothesis is rejected when $W_t$ stays outside the boundaries which is formed by the lines that connected the points $[k, \pm a(T - k)1/2]$ and $[k, \pm 3a(T - k)1/2]$. The corresponding $a$ for significance levels 1% and 5% are 0.948 and 1.143, respectively.

4 Errors in variables and test of stability

Errors in variables in regressors cause OLS estimators to be biased and inconsistent. Also it will lead to the heteroscedasticity. For example, in model 1, different measurement errors in the regressor $x_t$ in two subsets would cause unequal error variances in the regressor while different measurement errors in $y_t$ may lead to unequal error variances in the regression. Also, for trend regressors in economic time series data, the latest observations often have a comparatively large variance, which lead to unequal error variances for the whole time period. For the situation above, The OLS estimators $\hat{\beta}$ in different subsets of data might consist to different $\beta$ even if when the parameters are stable. This phenomenon will cause the empirical size to be higher than the theoretical significance level when the Chow test and the predictive Chow test are applied. Also the heteroscedasticity may cause the empirical sizes to be larger. These mean that we have more opportunities in rejecting the null hypothesis when it is true.

It is difficult to get analytical results when we have measurement errors. However, after making some assumptions, we may get comparatively simple results for the Chow test and the predictive Chow test. Assume think model 1 do not have intercept and there are measurement errors only in the last $T - T_1$ observations. We want to do the predictive Chow test to see if there is a structural break at time $T_1$. In an other words, we are interested in whether the parameter of the regressor is the same for all observations. Then the predictive Chow test statistics (15) become:

$$\frac{\sum_{i=1}^{T_1} (Y_t - \hat{\beta}_{T_1} X_t)^2 - \sum_{i=1}^{T_1} (Y_t - \hat{\beta}_{T_1} X_t)^2}{\sum_{i=1}^{T_1} (Y_t - \hat{\beta}_{T_1} X_t)^2/(T_1 - 1)}$$

where $\hat{\beta}_{T}$ and $\hat{\beta}_{T_1}$ and are OLS estimators for slope $\beta$ on $T, T_1$ observations. When $T_1$ is very large, the second subset becomes very small. Then the measurement error in last $T - T_1$ observations have little effect on OLS estimator $\hat{\beta}_{T}$ for the whole time period, indicating that $\hat{\beta}_{T}$ are approximately equal to the true regressor $\beta$. So the equation (25) can be approximated by:

$$\frac{\sum_{i=T_1+1}^{T} (Y_t - \hat{\beta}_{T} X_t)^2}{\sum_{i=1}^{T_1} (Y_t - \beta X_t)^2} \cdot \frac{T_1 - 1}{T - T_1}$$

(26)
If there are no measurement errors in last $T - T_1$ observations, (25) will follows an $F$ distribution with $T - T_1$ and $T_1 - 1$ degrees of freedom when the null hypothesis is true. However in fact we have measurement errors in second subset, then the OLS estimators are biased and the sizes will be larger than it should be.

5 Illustrations using simulated data

In this section we will illustrate how do the measurement errors affect the test of stability using the Chow test, the predictive Chow test and the CUSUM test. We will simulate samples and do these three types of tests on them. The significance level are 5\% and 1\%, and will be compared with the empirical sizes. In experiment 1 we will do all three types of tests. The intercept is set to 10 and the parameters of regressors is set to 1 in model 1. The regressors $x_t$ will be normally distributed with means 0 and variance equal to 4. For the Chow test and the predictive Chow test we split the observations into two groups, one with $T_1$ observations and the other $T - T_1$ observations. Because we have two type of errors, we let $\sigma^2_{u_1}$ and $\sigma^2_{u_2}$ denote the variances of the total error ($\sigma^2_{u_t} = \sigma^2_{q_t} + \sigma^2_{w_t}$), and $\sigma^2_{e_1}$ and $\sigma^2_{e_2}$ for error variances in the regressor. We vary these four variances to study the effect of measurement errors. The size of the sample is set to $T = 40$, a reasonable sample size for economic time series. We will also change the structural break point $T_1$, which might affect the results. In experiment 2 we study model 2. The intercept will be set to 10 and the two parameters of the regressors will be set to 1. Other parameters settings are the same as in experiment 1. In experiment 3 for model 3, three type of tests are used. The true values of $x_t$ are set to be a linear trend or a random walk with a drift $x_t = \alpha + x_{t-1} + \theta_t$. We assume the drift $\alpha = 1$ and the error $\theta_t$ is independently normally distributed with mean 0 and variance 1. Other settings are the same as in experiment 3. In this experiment we use the Chow test and the CUSUM test. In experiment 4 for model 1, the regressor is a linear trend $x = \{1, 2, \ldots, T\}$. Other settings are the same as in experiment 1. In experiment 5 for model 3, the true value $x = \{1, 2, \ldots, T\}$ is a linear trend or a random walk with a drift $x_t = \alpha + x_{t-1} + \theta_t$. We assume the drift $\alpha = 1$ and the error $\theta_t$ is independently normally distributed with mean 0 and variance 1. Other settings are the same as in experiment 3. In this experiment we use the Chow test and the CUSUM test. In experiment 6 for model 4, we have a instrumental variable regression $x_t = \delta_0 + \delta_1 z_t + \gamma_t$ where $z_t$ is an instrumental variable. Here we suppose $\delta_0$ is equal to 0 and $\delta_1$ is equal to 1. The error term $\gamma_t$ is independently normally distributed with means equal
to 0 and variance 1. Other settings are the same as in experiment 1. In this experiment we only do the Chow test.

5.1 Result of tests with normally distributed regressor
in model 1

As is mention in the section above, the true values of regressors are normally distributed with mean set to 0 and standard deviation to 2. The error variance in the structural equation ($\sigma_q^2$) is set to one. The relation between $R_{xy}^2$ and $R_{xy}^2$ is:

$$R_{xy}^2 = K_{xx} R_{xY}^2$$

(27)

where $R_{xy}^2$ is the squared correlation between $x_t$ and $Y_t$ and $R_{xY}^2$ is the squared correlation between $X_t Y_t$, $K_{xx}$ is the reliability ratio in (11). Again we refer to Fuller(1987) for more details. From (27) we can see that also $R^2$ is affected by the measurement error variance in the regressor($\sigma_q^2$).

If they are the same for the two groups, then the reliability ratio will be the same. Further more, if the error variance in $Y$ are equal because of homoscedasticity, then the empirical size would be close to the significance level. If, however, the error variances in $Y$ are not equal, we will have heteroscedasticity and also the structural break point $T_1$ would affect the test. So we systematically choose different error variances, different variances in the regressor and different time point $T_1$, then calculate the empirical sizes.

In Table 1, different significance level 1% and 5% are provided for three type of tests. From the first row for both levels we can see that without measurement error, the empirical sizes are similar in magnitudes to the significance levels stated for the three tests. When we have equal measurement errors in the regressors for all data, the empirical sizes coincides with the significance levels, meaning that the equal measurement errors does not affect these tests. In this situation the reliability ratio is 0.8 for different $T_1$. With unequal measurement errors in the regressors and unequal error variances, however, the sizes are out of control. In the third row, we assume that there are measurement error only in the second subset. We find that the sizes are too high compared with the significance level for the Chow test and the predictive Chow test. This becomes more serious as the second subset becomes smaller. When we add unequal measurement errors in the regressors and unequal error variances, which is presented in the last row, the empirical sizes are larger compared with the size in third row as the second subset becomes smaller. For the last two rows, the reliability ratio vary from 0.89 to 0.98. The results for the predictive Chow test are similar to those for the Chow test. For the CUSUM test, the empirical sizes in the first two situa-
tions (Without any error variances and with equal variances in the regressors) are basically the same as those stated. For the other two situations (row 3 and row 4), the empirical sizes are far too low, indicating that the CUSUM test faces more serious problems. From the simulations we can say the three type of tests can still be effective when we have same measurement errors variances in the regressors in model 1.

Then we want to simply see how does the sample sizes affect the three types of test. the Chow test is chosen in Table 2, which presents results for \( T = 40, T = 200 \) and \( T = 1000 \) sample sizes. We assume that there are no measurement errors in \( Y \) and fixed measurement errors in the regressors \( X \). Here the error variances in the regressor are set to \( \sigma^2_{\epsilon_1} = 0 \) and \( \sigma^2_{\epsilon_2} = 0.16 \). The reliability ratio is larger than 0.98 whenever we change \( T_1 \). The structural break point varies from \( T_1 = 0.5T \) to \( T_1 = 0.9T \). For this test with \( T = 40 \) observations, which is a reasonable number for econometric time series data, the empirical size is close to the significance level at a reasonable magnitude when the sample sizes are equal for two subsets. As the proportion of second subset becomes smaller, the empirical sizes increases rapidly. Compared with unequal sample sizes for two subsets, the increasing sample sizes influence the empirical size much more, which are almost 3 time for a sample size of 1000 observations as much as it for 200 observations.

**TABLE 1: Empirical sizes for three tests in presence of errors in variables with normally distributed regressor for model 1.**

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Error</th>
<th>( T_1=20 )</th>
<th>( T_1=24 )</th>
<th>( T_1=28 )</th>
<th>( T_1=32 )</th>
<th>( T=39 )</th>
<th>( T_1=32 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1,1,0,0</td>
<td>0.011</td>
<td>0.010</td>
<td>0.012</td>
<td>0.009</td>
<td>0.008</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>1,1,1,1</td>
<td>0.014</td>
<td>0.011</td>
<td>0.014</td>
<td>0.009</td>
<td>0.012</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>1,1,0,1</td>
<td>0.025</td>
<td>0.053</td>
<td>0.056</td>
<td>0.072</td>
<td>0.081</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>1,2,0,1</td>
<td>0.031</td>
<td>0.058</td>
<td>0.078</td>
<td>0.183</td>
<td>0.289</td>
<td>0.228</td>
</tr>
<tr>
<td>0.05</td>
<td>1,1,0,0</td>
<td>0.047</td>
<td>0.047</td>
<td>0.038</td>
<td>0.051</td>
<td>0.053</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>1,1,1,1</td>
<td>0.054</td>
<td>0.045</td>
<td>0.041</td>
<td>0.055</td>
<td>0.055</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>1,1,0,1</td>
<td>0.143</td>
<td>0.151</td>
<td>0.19</td>
<td>0.187</td>
<td>0.186</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>1,2,0,1</td>
<td>0.097</td>
<td>0.145</td>
<td>0.211</td>
<td>0.342</td>
<td>0.455</td>
<td>0.358</td>
</tr>
</tbody>
</table>
TABLE 2: Empirical sizes for the Chow tests for different sample sizes with normally distributed regressor for model 1.

<table>
<thead>
<tr>
<th>Sample size, T</th>
<th>Alpha</th>
<th>u1, u2, u3, u4</th>
<th>Chow T1=0.5T</th>
<th>Chow T1=0.7T</th>
<th>Chow T1=0.9T</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>0.01</td>
<td>1,1,0,0,16</td>
<td>0.013</td>
<td>0.016</td>
<td>0.019</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>1,1,0,0,16</td>
<td>0.016</td>
<td>0.019</td>
<td>0.024</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>1,1,0,0,16</td>
<td>0.057</td>
<td>0.057</td>
<td>0.036</td>
</tr>
<tr>
<td>40</td>
<td>0.05</td>
<td>1,1,0,0,16</td>
<td>0.049</td>
<td>0.062</td>
<td>0.075</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>1,1,0,0,16</td>
<td>0.062</td>
<td>0.084</td>
<td>0.074</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>1,1,0,0,16</td>
<td>0.161</td>
<td>0.175</td>
<td>0.129</td>
</tr>
</tbody>
</table>

5.2 Result of tests with normally distributed regressor in model 2

In this model, the part which is different from model (1) is that we add a new regressor $x_1$ which does not have measurement error. We intend to see whether a new regressor will affect three types of tests compared with model (1). Here, the true regressors follow i.i.d normal distribution with mean equal to 0 and variance equal to 4. Error variances ($\sigma^2_\epsilon$) are set to 1 in the structural equation.

Different combinations of error variances and significance levels are provided in Table 2. In this experiment we choose error variances $\sigma^2_u$, variances in the regressors $\sigma^2_\epsilon$, significance levels $\alpha$ and structural break point $T_1$ as we did in experiment 1. Without measurement errors, the empirical sizes are equal to the significance levels for both the Chow and the predictive Chow test when we take the random errors into consideration. The sizes for the CUSUM test are a bit smaller as it was for model (1). In the second row we can see that equal error variances for all data in one regressors make no influence on the empirical sizes. One of the reason is that the reliability ratios are the same for the two subsets and for the whole data set. Also the two subsets have homoscedasticity. Here the reliability ratio is 0.8. In last two rows, the sizes turn much higher when we have different errors in the regressor for two subsets. This is especially serious when the proportion of second subset is really small. In comparison with the results in model (1), however, the empirical sizes are smaller relatively, this turns out more and more obviously as the second subset becomes smaller. This phenomenon may owning to that a new normal distributed regressor mitigates the effect of measurement errors in test statistic for the Chow and the predictive Chow test, leading to a relative accurate result. The reliability ratios in last two rows are still between 0.89 ($T = 20$) to 0.98 ($T = 36$).
Table 4 illustrates how different sample sizes do affect the empirical sizes. The error variance $\sigma^2_u$, the variance in the regressor $\sigma^2_\varepsilon$, the significance level $\alpha$, the sample size $T$ and the structural break point $T_1$ are provided as in experiment 1. Here the reliability ratio is about 0.96 for the second subset. Note that we have no measurement error for regressor $x_1$ in the first subset ($\sigma^2_1 = 0$). From Table 4 we can see the empirical sizes in different sample sizes are similar in magnitudes with what was obtained for model (1). For small sample the sizes are appropriate (0.008 for $\alpha = 0.01$ and 0.048 for $\alpha = 0.05$) when the two subsets are equally large. As the sizes of the subsets are unequal, the empirical sizes increase almost 40%. Things are getting worse when we change the sample size. For 1% significance level, the empirical sizes are around 2% for 200 observations and they increase about 3 times (0.042) for 1000 observations. It means that whether we add a true regressor or not, varying the sample sizes including total sample size and size of subsets do affect the test results in the presence of measurement errors in the regressors. Was is so when the subsets are unequally large. Also we notice that most of the empirical sizes are a bit smaller, compared with that of model 1.

TABLE 3: Empirical sizes for three tests in presence of errors in variables with normally distributed regressor for model 2.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Error</th>
<th>T1=20</th>
<th>T1=24</th>
<th>T1=28</th>
<th>T1=32</th>
<th>T=36</th>
<th>T=39</th>
<th>T1=32</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>$u_1,u_2,e_1,e_2$</td>
<td>0.010</td>
<td>0.011</td>
<td>0.006</td>
<td>0.015</td>
<td>0.012</td>
<td>0.008</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>1,1,1,1</td>
<td>0.008</td>
<td>0.011</td>
<td>0.008</td>
<td>0.014</td>
<td>0.013</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>1,1,0,1</td>
<td>0.027</td>
<td>0.040</td>
<td>0.070</td>
<td>0.067</td>
<td>0.058</td>
<td>0.053</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>1,2,0,1</td>
<td>0.020</td>
<td>0.061</td>
<td>0.122</td>
<td>0.224</td>
<td>0.264</td>
<td>0.220</td>
<td>0.004</td>
</tr>
<tr>
<td>0.05</td>
<td>1,1,0,0</td>
<td>0.046</td>
<td>0.047</td>
<td>0.04</td>
<td>0.049</td>
<td>0.052</td>
<td>0.051</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>1,1,1,1</td>
<td>0.049</td>
<td>0.040</td>
<td>0.051</td>
<td>0.052</td>
<td>0.046</td>
<td>0.048</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>1,1,0,1</td>
<td>0.109</td>
<td>0.170</td>
<td>0.173</td>
<td>0.205</td>
<td>0.254</td>
<td>0.156</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>1,2,0,1</td>
<td>0.095</td>
<td>0.158</td>
<td>0.263</td>
<td>0.387</td>
<td>0.423</td>
<td>0.361</td>
<td>0.007</td>
</tr>
</tbody>
</table>
TABLE 4: Empirical sizes for the Chow test in presence for different sample sizes with normally distributed regressor for model 2.

<table>
<thead>
<tr>
<th>Sample size, T</th>
<th>Size %</th>
<th>u1,u2,e1,e2</th>
<th>Chow T1=0.5T</th>
<th>Chow T1=0.7T</th>
<th>Chow T1=0.9T</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>1</td>
<td>1,1,0,0.16</td>
<td>0.008</td>
<td>0.012</td>
<td>0.014</td>
</tr>
<tr>
<td>200</td>
<td>1</td>
<td>1,1,0,0.16</td>
<td>0.015</td>
<td>0.023</td>
<td>0.029</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>1,1,0,0.16</td>
<td>0.042</td>
<td>0.053</td>
<td>0.032</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>1,1,0,0.16</td>
<td>0.048</td>
<td>0.057</td>
<td>0.069</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>1,1,0,0.16</td>
<td>0.059</td>
<td>0.079</td>
<td>0.081</td>
</tr>
<tr>
<td>1000</td>
<td>5</td>
<td>1,1,0,0.16</td>
<td>0.147</td>
<td>0.163</td>
<td>0.112</td>
</tr>
</tbody>
</table>

5.3 Result of tests with normally distributed regressor in model 3

Assume in model: \( y_t = \beta_0 + \beta_1 x_t + q_t \), the true regressor value \( \{x_t\} \) are independently normally distributed with mean 0 and variance 4. The error variance \( \{q_t\} \) is set to be a random walk, which is \( q_t = q_{t-1} + p_t \), where \( p_t \) is independently normally distributed. In this experiment the variance of errors \( p_t \) is set to 1. So after taking the first difference of the model, the new model becomes model 3. Different combinations of error variances, structural change points are chosen and the empirical sizes are presented in Table 5, which gives the result for 1% and 5% significance levels. For the Chow test and the predictive Chow test, when there are no measurement errors, the empirical sizes are of the same magnitudes as the stated levels, while the sizes for the CUSUM test are smaller, comparatively. In the second row, we can see that the sizes for the Chow test and the predictive Chow test are still almost the same as the significance levels when we add measurement errors in the regressors with equal variances for two subsets. For the CUSUM test, however, the sizes are too small. For 1% significance level, the size is approximate 0, indicating a loss of power of this test. In the last row for each level, we have measurement error only in second subset. The empirical sizes are too large compared with stated levels for the Chow and the predictive Chow test, while the sizes for the CUSUM test are too low. Also the effects of measurement errors are relatively larger for 1% significance level than for 5%.

Let us briefly make a comparison between this model and model 1. As we give above, model 3 can be derived from model 1 by taking first difference when the error term in model 1 are random walk. It is interesting to find that the sizes in third row for each level in Table 5 are relatively larger than
it in third row in Table 1 for the Chow test and the predictive Chow test, while the size for the CUSUM test is smaller. This may owing to that the measurement error in model 3 is larger than in model 1. Suppose we have equal measurement errors $\varepsilon_t$ for $x_t$ in both model 1 and model 3. Then the series $\{\nabla x_t\}$ will have measurement error $\{\varepsilon_t - \varepsilon_{t-1}\}$. The error variances in the regressor $\nabla x_t$ will be $2\sigma^2$, which will affect the test much more. Now assume both model 1 and 3 don’t have measurement error in dependent variable $y$, and only have measurement error in the regressor in the second subset. The empirical sizes are shown in Table 6. We can see that for the Chow test the sizes in model 1 are of the same magnitudes as it was obtained for model 3 when the error variance of regressor in model 1 is $\sqrt{2}$ time the same as it in model 3 for each levels. This means that taking first difference is not a good choice to do structural change tests when we have measurement errors in the normally distributed regressor.

**TABLE 5: Empirical sizes for three tests in presence for different sample sizes with normally distributed regressor for model 3.**

<table>
<thead>
<tr>
<th>Alpha</th>
<th>u1,u2,u1,e2</th>
<th>T1=20</th>
<th>T1=24</th>
<th>T1=28</th>
<th>T1=32</th>
<th>T=39</th>
<th>T=32</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1,1,0,0</td>
<td>0.009</td>
<td>0.011</td>
<td>0.010</td>
<td>0.009</td>
<td>0.009</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>1,1,1,1</td>
<td>0.011</td>
<td>0.008</td>
<td>0.009</td>
<td>0.008</td>
<td>0.011</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>1,1,0,1</td>
<td>0.053</td>
<td>0.082</td>
<td>0.100</td>
<td>0.132</td>
<td>0.152</td>
<td>0.058</td>
</tr>
<tr>
<td>0.05</td>
<td>1,1,0,0</td>
<td>0.051</td>
<td>0.052</td>
<td>0.053</td>
<td>0.051</td>
<td>0.054</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>1,1,1,1</td>
<td>0.054</td>
<td>0.041</td>
<td>0.040</td>
<td>0.034</td>
<td>0.047</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>1,1,0,1</td>
<td>0.165</td>
<td>0.198</td>
<td>0.239</td>
<td>0.249</td>
<td>0.284</td>
<td>0.147</td>
</tr>
</tbody>
</table>

**TABLE 6: Comparison of empirical sizes for the the Chow test in presence of error in the regressor with normally distributed regressor for model 3.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Alpha</th>
<th>Variance of $e^2$</th>
<th>Chow $T1=20$</th>
<th>Chow $T1=24$</th>
<th>Chow $T1=28$</th>
<th>Chow $T1=32$</th>
<th>Chow $T1=36$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>$\sqrt{2}$</td>
<td>0.055</td>
<td>0.081</td>
<td>0.108</td>
<td>0.127</td>
<td>0.134</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>1</td>
<td>0.056</td>
<td>0.079</td>
<td>0.102</td>
<td>0.125</td>
<td>0.137</td>
</tr>
<tr>
<td>1</td>
<td>$\sqrt{2}/2$</td>
<td>0.022</td>
<td>0.038</td>
<td>0.043</td>
<td>0.051</td>
<td>0.061</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$1/2$</td>
<td>0.022</td>
<td>0.034</td>
<td>0.045</td>
<td>0.048</td>
<td>0.068</td>
<td></td>
</tr>
</tbody>
</table>

### 5.4 Result of tests with trending regressor in model 1

In this experiment, the total sample size is set to $T = 40$, which is a reasonable number for econometric time series data. And $x = 1,2,\ldots,T$ is a
trending regressor. In Table 7, three types of test are presented: the Chow test, the predictive Chow test and the CUSUM test. The significance levels \( \alpha \) and structural break point \( T_1 \) are stated as we did in experiment 1 and 2. However here we provide different combinations of measurement errors. From the first two rows for each significance level we can see that the empirical sizes for the three tests are close to those stated, when there is no measurement error in the regressors. The empirical sizes are larger than the significance levels, when we let error variances in the regressors to be the same in two subsets. For the Chow test, the empirical sizes increase as the measurement error in the regressor increase, given that the error variances \( u_t \) are fixed. For given error variances in the regressor, however, the empirical sizes decrease as the error variances increase. Also varying the structural break point \( T_1 \) does not influence the sizes. For the predictive Chow test, the sizes are of same magnitudes as stated levels, meaning that equal variances for \( u_t \) and equal error variances in the regressors do not influence the tests. However the the CUSUM test preforms badly. The empirical sizes are too large when the measurement error in the regressor goes up, even the size reach 50 times as large as the significance level 1%. Then we set the measurement error variance in the first subset to zero (\( \sigma^2_{e_1} = 0 \)). For the Chow test, the larger the measurement errors in the regressor, the larger empirical sizes are. And it become more and more obvious especially when the second subset is rather small. For the predictive Chow test, the sizes are too high, indicating that we have more chance to reject a true null hypothesis. For the CUSUM test, however, the sizes are too small, indicating a loss of power of this test.
5.5 Result of tests with trending regressor in model 3

Assume in model: \( y_t = \beta_0 + \beta_1 x_t + q_t \), that the true regressor \( \{x_t\} \) is a trending regressor according to \( x = \{1, 2, \ldots, 40\} \). The error \( \{q_t\} \) is set to be a random walk, which is \( q_t = q_{t-1} + p_t \), where \( p_t \) is independently normally distributed. In this experiment the variance of errors \( p_t \) is set to be 1. So after taking the first difference, the new model becomes model 3 as we had in section 5.3. The new regressor turns out to be \( \nabla x = \{1, 1, \ldots, 1\} \) after taking first difference. It is obvious that the variances of \( \nabla x \) is 0. So the reliability ratio is \( \frac{\text{var}(x)}{\text{var}(x) + \text{var}(\varepsilon)} \) = 0, where \( \varepsilon \) stands for the measurement error for the regressor \( \nabla x \). Different combinations of error variances, structural change points are provided and the empirical sizes are presented in Table 8, which gives the results for 1% and 5% significance levels. For the Chow test, from the first row for each level we note that the empirical sizes are far to
low compared with the significance levels. This may because the variance of the regressor \( x \) is equal to zero. For the CUSUM test, however, the size is a bit lower than the given one. In the second row, when we add same measurement errors to the regressor the sizes are equal to the magnitudes of the significance levels while the sizes for the CUSUM test are still too small. In the third row we can see that the change of the variances in the regressor for the whole sample does not influence the sizes: the sizes are close to the stated ones. In the last three rows, we assume that the measurements errors in the regressor for each group are unequal. For the Chow test, the sizes are still equal in magnitude to the significance levels, indicating that the unequal error variances do not affect the sizes. Also varying the sample sizes for each subset do not change the empirical sizes.

Now let us add some variances for the regressor \( \nabla x_t \) so the reliability ratio will not be zero. Suppose \( x \) is random walk with a drift: \( x_t = \alpha + x_{t-1} + \theta_t \) where \( \alpha = 1 \) and \( \theta_t \) are independent normally distributed with mean 0 and variance 1. Other assumption are the same as above. When we take the first difference, \( \nabla x_t = \alpha + \theta_t \) and the reliability ratio \( \sigma^2_\theta / (\sigma^2_\theta + \sigma^2_\varepsilon) \) will not be zero. The result of the simulation are provided in Table 9. Here we notice that for the Chow test the empirical sizes are equal to the magnitudes of the significance levels and for the CUSUM test the empirical sizes are somewhat smaller whether there are measurement errors or not. So we can conclude that the Chow test is stable for this situation.

Also we note that model 3 with a random walk regressor is similar to model 1. In model 1, we have: Structural equation:

\[
y_t = \beta_1 + \beta_2 x_t + q_t, \text{ where } q_t \text{ is } \text{IN}(0, \sigma^2_q), t = 1, \ldots, T \tag{28}
\]

Here we assume \( x_t \) to have measurement errors. For model 3 we have:

\[
\nabla y_t = \beta \nabla x_t + p_t, \text{ where } p_t = q_t - q_{t-1}, \text{ and } p_t \text{ is } \text{IN}(0, \sigma^2_q), t = 1, \ldots T \tag{29}
\]

Here too we assume \( x_t \) to have measurement errors. From section 2 we know that in (28) and (29), the regressors \( x_t \) and \( \nabla x_t \) are normally distributed. If we ignore \( \beta_1 \), then the only difference between the two models is the measurement error part. In (28), the regressor \( x_t \) has independently normally distributed measurement error \( \varepsilon_t \) while in (29), the regressor \( \nabla x_t \) do not. In (29), while \( x_t \) have measurement errors \( \varepsilon_t \), the regressors \( \nabla X_t = X_t - X_{t-1} = (x_t + \varepsilon_t) - (x_{t-1} + \varepsilon_{t-1}) = \nabla x_t + \varepsilon_t - \varepsilon_{t-1} \). So the measurement error part will be \{\nabla \varepsilon_t\}, indicating the measurement errors for the regressor are correlated.
TABLE 8: Empirical sizes for two tests in presence of errors in variables with trending regressor for model 3.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Error</th>
<th>Chow</th>
<th>CUSUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>e1,e2</td>
<td>T1=20</td>
<td>T1=24</td>
</tr>
<tr>
<td>1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,0</td>
<td>0.002</td>
<td>0.001</td>
<td>0.0025</td>
</tr>
<tr>
<td>1.1</td>
<td>0.0095</td>
<td>0.011</td>
<td>0.0105</td>
</tr>
<tr>
<td>2.2</td>
<td>0.0125</td>
<td>0.0097</td>
<td>0.011</td>
</tr>
<tr>
<td>0,1</td>
<td>0.013</td>
<td>0.0095</td>
<td>0.0125</td>
</tr>
<tr>
<td>0,2</td>
<td>0.0085</td>
<td>0.012</td>
<td>0.0105</td>
</tr>
<tr>
<td>1,2</td>
<td>0.0110</td>
<td>0.0092</td>
<td>0.0102</td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,0</td>
<td>0.009</td>
<td>0.0135</td>
<td>0.013</td>
</tr>
<tr>
<td>1,1</td>
<td>0.051</td>
<td>0.047</td>
<td>0.0485</td>
</tr>
<tr>
<td>2,2</td>
<td>0.049</td>
<td>0.0475</td>
<td>0.0495</td>
</tr>
<tr>
<td>0,1</td>
<td>0.053</td>
<td>0.0465</td>
<td>0.055</td>
</tr>
<tr>
<td>0,2</td>
<td>0.0582</td>
<td>0.0535</td>
<td>0.0655</td>
</tr>
<tr>
<td>1,2</td>
<td>0.049</td>
<td>0.052</td>
<td>0.039</td>
</tr>
</tbody>
</table>
### Table 9: Empirical sizes for two tests in presence of errors in variables with random walk with a drift regressor for model 3.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Error</th>
<th>Chow (T1=20)</th>
<th>Chow (T1=24)</th>
<th>Chow (T1=28)</th>
<th>Chow (T1=32)</th>
<th>CUSUM (T1=32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>e1,e2</td>
<td>0.012</td>
<td>0.011</td>
<td>0.0092</td>
<td>0.0125</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0096</td>
<td>0.012</td>
<td>0.0115</td>
<td>0.0127</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0095</td>
<td>0.0089</td>
<td>0.0096</td>
<td>0.011</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.011</td>
<td>0.0115</td>
<td>0.0095</td>
<td>0.0076</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0092</td>
<td>0.012</td>
<td>0.0105</td>
<td>0.009</td>
<td>0.0115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0105</td>
<td>0.0092</td>
<td>0.0102</td>
<td>0.0095</td>
<td>0.0110</td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td>0.049</td>
<td>0.0465</td>
<td>0.050</td>
<td>0.055</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.052</td>
<td>0.048</td>
<td>0.0492</td>
<td>0.056</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.048</td>
<td>0.049</td>
<td>0.0495</td>
<td>0.049</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.057</td>
<td>0.055</td>
<td>0.057</td>
<td>0.048</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.052</td>
<td>0.0515</td>
<td>0.0555</td>
<td>0.055</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.046</td>
<td>0.053</td>
<td>0.048</td>
<td>0.053</td>
<td>0.051</td>
</tr>
</tbody>
</table>

#### 5.6 Result of tests for instrumental variable method with normally regressor in model 4

In this model we assume that the regressor $x_t$ is normally distributed with mean 0 and variance 4. The variance of $q_t$ is set to 1. For the structural equation, the intercept is set to 10 and the slope to one for all observations while for the instrumental variable equation, we assume $\delta_0$ is equal to 0 and $\delta_1$ is equal to 1. It is easy to find that the structural model is the same as model 1. Also the variances of error in equation $\sigma^2_\gamma$ is set to 1. Note that whether we change the error variances for the regressor ($\sigma^2_{x1}$ and $\sigma^2_{x2}$), the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased and consistent.

Different combinations of error variances in the regressors are chosen and the empirical sizes for the Chow test are provided in Table 10. From the first row we can see that without measurement errors the sizes are equal in magnitudes to the given levels. Adding equal error variances in the regressor for two subsets do not affect the empirical sizes. In last two rows, we assume that only the second subsets have measurement errors in the regressor. We find that the sizes are too large especially when the second subsets are relatively small.
small. Also the sizes are larger as the variance of measurement errors in the regressor becomes larger and this is obvious when there are fewer observations in the second subset. In this simulation we can see that even when we have unbiased estimators for the parameters, the measurement errors still affect the structural change test. The instrumental variables method does not make a good improvement for the Chow test because this test is sensitive for heteroscedasticity. See Schmidt (1977) and Toyoda (1974) for more details.

**TABLE 10:** Empirical sizes for the Chow tests in presence of errors in variables with normally distributed regressor for model 4.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Error</th>
<th>T1=20</th>
<th>T1=24</th>
<th>T1=28</th>
<th>T1=32</th>
<th>T1=36</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1,1,0,0</td>
<td>0.01</td>
<td>0.0095</td>
<td>0.009</td>
<td>0.0095</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>1,1,1,1</td>
<td>0.01</td>
<td>0.012</td>
<td>0.013</td>
<td>0.01</td>
<td>0.0125</td>
</tr>
<tr>
<td></td>
<td>1,1,0,1</td>
<td>0.0355</td>
<td>0.0415</td>
<td>0.0555</td>
<td>0.053</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>1,1,0,2</td>
<td>0.064</td>
<td>0.0755</td>
<td>0.1155</td>
<td>0.135</td>
<td>0.142</td>
</tr>
<tr>
<td>0.05</td>
<td>1,1,0,0</td>
<td>0.0475</td>
<td>0.0405</td>
<td>0.0415</td>
<td>0.043</td>
<td>0.0445</td>
</tr>
<tr>
<td></td>
<td>1,1,1,1</td>
<td>0.0465</td>
<td>0.0445</td>
<td>0.0455</td>
<td>0.048</td>
<td>0.0515</td>
</tr>
<tr>
<td></td>
<td>1,1,0,1</td>
<td>0.1105</td>
<td>0.1145</td>
<td>0.123</td>
<td>0.1435</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>1,1,0,2</td>
<td>0.165</td>
<td>0.204</td>
<td>0.2205</td>
<td>0.237</td>
<td>0.2535</td>
</tr>
</tbody>
</table>

### 6 Conclusion

Testing for structural change in the presence of errors in variables may be risky for most cases of econometrics analysis and time series analysis. In this paper we have studied the empirical sizes for the Chow test, the predictive Chow test and the CUSUM test under different assumptions of errors in variables. Also we have studied the behavior of four types of models: a one regressor model, a two regressors model, a first difference model and an instrumental variable model. For the first difference model with trending regressor or random walk regressor, the Chow test and the CUSUM test perform well. The empirical sizes are the same as the magnitudes of the stated significance levels. In other situations, however, the empirical sizes for the Chow test and the predictive Chow test are much too high compared with the significance levels when we have different error structures in the two sub-
sets of data. For the CUSUM test, the corresponding empirical sizes are too low, indicating a loss of power of this test. These phenomena show that the test of structural change are sensitive when we have errors in variables. In general we have thought one could have different probability limits for parameter estimates in two subset of data compared with those for the data set as a whole contributed to the overestimated significance levels. However the instrumental variables model proves this to be false. In this paper it clearly demonstrate that even when we have same unbiased and consistent estimators of parameters for two subsets, the test of structural change in presence of errors in variables could still be risky. One reason for overestimating the significance levels could be heteroscedasticity which is caused by the measurement errors. For the Chow test and the predictive Chow test, the significance levels are often overestimated while for the CUSUM test they are overestimated or underestimated in different situations. This influence of heteroscedasticity for structural change test could be ignored when we take first difference for trending regressors or random walk regressors. So when there is an error in variables situation we should be careful in explain the result of the test of structural change. This is in line with Jonsson(1997)

In this paper all the models contain one or two regressors. More work can be done for multiple regressions. Also the first difference method with trending regressors and random walk regressors is a good way of dealing with structural change in presence of error in variable. This type of model is worth to study further.
References


