BAYESIAN AND FREQUENTIST HYPOTHESIS TESTS OF HETEROSCEDASTICITY

WeiWei Feng
1981-09-01
Abstract

The homoscedasticity assumption is important for the classical linear regression. This assumption is often violated in time series data, cross section data or panel data. In order to address this issue, the generalized linear regression or feasible generalized linear regression is suggested. However, applying those methods requires either knowledge of structure of variances or estimation of this structure. Before running into such a complex process, testing the heteroscedasticity is not only important but also necessary. In this paper, we used the Monte Carlo simulation to compare both the size and the power of Bayesian hypothesis test with the frequentist hypothesis tests for heteroscedasticity. The Bayesian hypothesis in this case is unpractical and less effective comparing with frequentist hypothesis tests. However, the Bayesian heteroscedasticity test could be possibly improved and the Bayesian heteroscedasticity model could be applied in other situations such as estimations of parameters or the structure of variances.

Keywords: Bayesian Hypothesis test, the White test, the Breusch-Pagan Test, the Koenkar-Basset Test, the Gibbs sampling, the Metropolis-hasting sampler, the marginal likelihood simulation.
Table of Contents

Introduction ........................................................................................................................................... 1

The frequentist hypothesis tests ............................................................................................................. 3

The White test ........................................................................................................................................ 3
The Breusch-Pagan Test ............................................................................................................................ 4
The Koenker-Basset Test .......................................................................................................................... 6

The Bayesian Hypothesis test ................................................................................................................... 7

The heteroscedasticity Bayesian model $M_{\text{heter}}$ and the homoscedasticity $M_{\text{hom}}$ .......................... 9
  Heteroscedasticity model $M_{\text{heter}}$ .................................................................................................. 9
  Homoscedasticity model $M_{\text{hom}}$ .............................................................................................. 13

Bayes factor simulation ............................................................................................................................ 14
The tests for convergence .......................................................................................................................... 19

The size and the power ............................................................................................................................. 21

In the frequentist hypothesis test framework .......................................................................................... 21
In the Bayesian hypothesis test framework .............................................................................................. 21

Monte Carlo Simulation ............................................................................................................................ 23

Data generating process .......................................................................................................................... 23
Parameters for the prior distributions in Bayesian models ........................................................................ 25
Starting points in Bayesian Markov Chain Monte Carlo simulation ........................................................ 29
Other parameters in simulation study ....................................................................................................... 29

Monte Carlo Results ............................................................................................................................... 30

Conclusion and Discussion ...................................................................................................................... 33

References ............................................................................................................................................... 34

Appendix ............................................................................................................................................... 35
Introduction

When testing the assumption of homoscedasticity in the classical linear regression, a number of the literatures can be gathered within the classical statistic. White (1980) is the most cited article, the well-known Breusch and Pagan (1979) and an improved test proposed by Koenker (1981), Godfrey (1996) and Koenker and Bassett (1982). On the other hand, there are few literatures about the Bayesian hypothesis test for heteroscedasticity. Richard Startz (2012) constructed Bayesian heteroscedasticity-Robust Standard Errors. Differing from his model, a hierarchical Bayesian model is constructed in order to conduct a Bayesian hypothesis test for heteroscedasticity. Both the size and the power will be compared with three common used frequentist hypothesis tests: the White’s test (White, 1980), the Breusch-Pagan test (Breusch and Pagan, 1979), and the Koenkar-Basset test (Koenker and Bassett, 1982).

In the classical linear regression model, one of the important assumptions is homoscedasticity and nonautocorrelation. It means each disturbance, \( \varepsilon_i \), has the same finite variance, \( \sigma^2 \), and is uncorrelated with any other disturbances. This assumption limits the generality of the model and is often violated in the real world.

Heteroscedasticity means that disturbances have different variances, which often occurs in time-series data, cross-section data or panel data. In this paper, nonautocorrelation is assumed for simplicity. The linear regression model with heteroscedasticity can be expressed as follow:

\[
\begin{align*}
    y &= X\beta + \varepsilon \\
    E[\varepsilon|X] &= 0 \\
    E[\varepsilon'\varepsilon|X] &= \sigma^2 \Lambda = \Sigma \\
    \sigma^2 \Lambda &= \sigma^2 \begin{pmatrix}
        \lambda_1 & \cdots & 0 \\
        \vdots & \ddots & \vdots \\
        0 & \cdots & \lambda_n
    \end{pmatrix} = \begin{pmatrix}
        \sigma_1^2 & \cdots & 0 \\
        \vdots & \ddots & \vdots \\
        0 & \cdots & \sigma_n^2
    \end{pmatrix}
\end{align*}
\]

(1)
Where $y$ is a vector of dependent variables, $X$ is a $k \times n$ matrix of $k$ independent variables including the constant term, $n$ is number of observations, and $\beta$ is a vector of coefficients. While $\epsilon$ is a vector of disturbances and assumed normally distributed in ordinary situations. $\Sigma$ is a variance and covariance matrix. $\Lambda$ is a positive definite matrix where $\Lambda = I$ means homoscedasticity.

Even with heteroscedasticity, under the assumption $E[\epsilon|X] = 0$ and the normality assumption of the disturbance, the least squares estimator $\hat{\beta} = (X'X)^{-1} X' y$ is still unbiased, consistent and asymptotically normally distributed. However, the estimates are not BLUE (best linear unbiased estimator) as showed in the Gauss-Markov Theorem. The variance estimator $s^2 (X'X)^{-1}$ is incorrect and leads to unreliable confidence interval estimation and hypothesis tests. Estimation of the asymptotic covariance matrix then could be based on

$$\text{var}(\hat{\beta}|X) = (X'X)^{-1} X' (s^2 \Lambda) X (X'X)^{-1}$$

(2)

If $\Lambda$’s structure is known, the model could be transformed to a homoscedasticity model and ordinary least square (OLS) is still BLUE. In a more common case, the structure is impossible to know. So $\hat{\Lambda}$ is estimated according to a different approach called the method of feasible generalize least squares (FGLS). (Green, 2012)

Before rushing into such a complex process, it is best to test whether heteroscedasticity exists in the data. This paper mainly focuses on comparing both the size and the power between frequentist heteroscedasticity tests and the Bayesian hypothesis test.
The frequentist hypothesis tests

The three common used classical tests are the White test (White, 1980), the Breusch-Pagan test (Breusch and Pagan, 1979), and the Koenkar-Basset test (Koenker and Bassett, 1982).

They are based on the idea that OLS is a consistent estimator of $\beta$ even in the presence of heteroscedasticity. In this case, the OLS residuals will preserve information, although not perfectly, about the heteroscedasticity. Therefore, the tests are applied to the OLS residuals in order to detect heteroscedasticity in the data.

The above-mentioned three tests are all derived from the Lagrange multiplier test (LM). The LM test assumes a properly scaled Lagrange multiplier has an asymptotically normal distribution. The LM test statistic is a quadratic form of the Lagrange multiplier. The test statistic, under the null hypothesis, is then chi-squared distributed.

The White test

The White test for heteroscedasticity has a basic idea: if the model is homoscedastic, then the disturbances are randomly distributed and have no relationship with any independent variables and their combinations such as squared form or a cross product form. The test statistic is derived based on the data with no heteroscedasticity, and then the least squares estimator of variance gives a consistent estimator of $\text{var}[b|X]$ as in the equation (2) (Green, 2012, p.315).

Null hypothesis $H_0$: Homoscedasticity ($\sigma_i^2 = \sigma^2$ for all $i$ in model (1))
Alternative hypothesis $H_1$: heteroscedasticity (not all $\sigma_i^2$ are equal)

The testing process can be simplified as follows:

1) Regress the dependent variable $y$ with all the independent variables $X$. 
2) Compute the OLS residuals, \( e_1, e_2, \ldots, e_n \). \( e \) is the OLS estimator of \( \varepsilon \) in the model (1).

3) Regress \( e_i^2 \) against a constant, all independent variables, their squared form, and their cross products.

4) Compute \( R^2 \) from this “auxiliary regression” in Step 3.

5) Compare \( nR^2 \) to the critical value from the Chi-squared distribution with \( p \) degrees of freedom.

The test statistic is \( nR^2 \sim \chi^2_{p-1} \), where \( p \) is the number of the regressor in this “auxiliary regression” including the constant in step 3 and \( n \) is the number of observations.

The White test is very general and the specific assumption about the nature of the heteroscedasticity is not necessary to make. However, because of its generality, it has a serious shortcoming. The test may reveal heteroscedasticity, but does not detect the reason for it. This causes some trouble when applying the generalized least squares (GLS) method, for which we need to know the structure of the variance. At the same time, the degrees of freedom grow rapidly with the number of independent variables. Hence, it may be just appropriate for the relatively large sample size.

One modification can be done in order to lower the degrees of freedom. In step 3, we regress \( e^2 \) against \( \hat{\gamma} \) and \( \hat{\gamma}^2 \) instead, where \( \hat{\gamma} = Xb \), \( b \) is the OLS estimator of the parameters in model (1). In this way, the regressors actually include all independent variables, their squared form and their cross products. In this method, the degrees of freedom decrease to 2 and the test statistic is \( nR^2 \sim \chi^2(2) \). This method is used instead of the original White test in our study.

**The Breusch-Pagan Test**

The Breusch-Pagan Test is derived from an LM test with the hypothesis that

\[
\sigma_i^2 = \sigma^2 f(Z\alpha),
\]

where \( Z \) is a \( p \times n \) matrix of special combination of independent variables, squared form of independent variables and the cross products of independent variables that is suspected to cause the heteroscedasticity, \( p \) is the number of regressors.
including a constant and n is the number of observations, and f is a function. The model is homoscedastic if \( \alpha = 0 \), \( \alpha \) is a vector of parameters in the equation. The test can be carried out with a simple regression.

Null hypothesis \( H_0 \): Homoscedasticity (\( \alpha = 0 \))

Alternative hypothesis \( H_1 \): Heteroscedasticity (\( \alpha \neq 0 \))

1) Regress the dependent variable \( y \) with all the independent variables \( X \).

2) Compute the OLS residuals, \( e_1, e_2, ..., e_n \). \( e \) is the vector with all OLS estimator of disturbance \( e_i \) in the model (1) and estimates the variance of disturbance as \( e'e / n \).

3) Regress \( e^2 / (e'e / n) \) against \( Z \). \( e^2 \) is the vector of all the squared \( e_i \) and compute the ESS (explained sum of square) of the regression. The test statistic ESS is asymptotically distributed as \( \chi^2_{p-1} \), \( p \) is the number of regressors in \( Z \) including a constant. This test statistic ESS can be directly computed with the matrix

form \( LM = \frac{1}{2} [g'Z(Z'Z)^{-1}Zg] \); \( g \) is the vector of observations of \( g_i = e_i^2 / (e'e / n) - 1 \).

(Breusch and Pagan, 1979)

4) Compare the value of the test statistic from step 4 to the critical value from the Chi-squared distribution with \( p-1 \) degrees of freedom. Homoscedasticity is rejected if the value exceeds the critical value.

The Breusch-Pagan Lagrange multiplier test is claimed to be sensitive to the normality assumption. Violation of this assumption does not guarantee that the variance of \( \varepsilon_i^2 \) is equal to \( 2\sigma^4 \) under homoscedasticity. In this situation, the test statistic is incorrect.

(Green, 2012, p.276)
The Koenker-Basset Test

Koenker and Basset (1982) has the same $H_0$ and $H_1$ and almost the same process as the Breusch-Pagan test.

Null hypothesis $H_0$: Homoscedasticity ($\alpha = 0$)

Alternative hypothesis $H_1$: Heteroscedasticity ($\alpha \neq 0$)

However, it suggests a more robust estimator of the variance of $\varepsilon_i^2$:

$V = \frac{1}{n} \sum_{i=1}^{n} (\varepsilon_i^2 - \varepsilon' \bar{\varepsilon})^2$.

With this change the test statistic in Koenker and Basset test becomes

$LM = \left[ \frac{1}{V} (\varepsilon'^2 - \varepsilon' \bar{\varepsilon})' Z(Z'Z)^{-1} Z' (\varepsilon'^2 - \varepsilon' \bar{\varepsilon}) \right]$, which is asymptotically distributed as $\chi^2_{p-1}$.

Where $p$ is the number of regressors in $Z$ (has the same structure as in the Breusch-Pagan test) including a constant (Green, 2012). Other steps are the same as the Breusch-Pagan test.

Under normality, the Koenker-Basset test has the same limiting distribution as the Breusch-Pagan statistic, but violating the normality, it can be a more powerful test. And if $Z$ has the same structure as the regressor in the white test, the Koenker-Basset test is algebraically the same as the White test (Waldman, 1983).

In the simulation part of the Breusch-Pagan test and the Koenker-Basset test, $Z$ is some partition of combination of independent variables, square of independent variables and cross product of independent variables.
The Bayesian Hypothesis test

In the frequentist hypothesis tests framework, there are two hypotheses: the Null hypothesis $H_0$ and the alternative hypothesis $H_1$. A decision is made based on the collected data. The Null hypothesis is either rejected in favor of the alternative hypothesis, or accepted. In Bayesian hypothesis testing, there could be more than two hypotheses.

The Bayesian hypothesis test will be explained in the way that only two of the hypothesis will be chosen between.

The model can be generally expressed as follows:

\[ p(y|\theta_{M_i},M_i) \text{-- A data likelihood distribution} \]

\[ p(\theta_{M_i}|M_i) \text{-- A prior distribution under the model } M_i \]

\[ p(M_i) \text{-- A prior probability that the model is "correct"} \]

(3)

The models can be different on the data likelihood distribution, a prior distribution or both. In this paper, the heteroscedasticity model and the homoscedasticity model are different in both the data distribution and prior distributions.

Before explaining the Bayesian hypothesis test, the loss function is introduced first.

\[ L(M_i,M_j) = 0 \text{ when model } M_i \text{ is "correct"} \]

\[ L(M_i,M_j) > 0 \text{ when model } M_i \text{ is "wrong", } i \neq j \]

The posterior expected loss from choosing model $i$ based on the data $y^o$ is

\[ E[L(M_i,M_j)] = \sum_j L(M_i,M_j)p(M_j|y^o) \]

The model with the smallest posterior expected loss would be chosen. $y^o$ means the collected data.

In a situation with only two hypotheses, the expected lost for model $M_i$:

\[ E[L(M_i,M_j)] = L(M_i,M_j)p(M_j|y^o) \]

The expected loss for model $M_j = E[L(M_j,M_i)] = L(M_j,M_i)p(M_i|y^o), \ i,j = 1,2 \ i \neq j$. 
If $L(M_i, M_j)p(M_j | y^o) < L(M_j, M_i)p(M_i | y^o)$, the model $M_i$ will be chosen over model $M_j$ based on the data information $y^o$.

In this inequality, the $p(M_j | y^o)$ is the posterior model probability which follows Bayes rules

$$P(M_i | y) = \frac{p(y | M_i)p(M_i)}{\sum_j p(y | M_j)p(M_j)} \tag{4}$$

$$p(y | M_i) = \int p(y | \theta_{M_i})p(\theta_{M_i} | M_i)d\theta \tag{5}$$

By putting the equality (4) into the inequality, because the denominators are the same for both models, we get:

$$B_{ij} = \frac{p(y^o | M_i)}{p(y^o | M_j)} > \frac{L(M_i, M_j)p(M_j)}{L(M_j, M_i)p(M_i)} \tag{6}$$

$B_{ij}$ is a Bayes factor. When it is larger than the right-hand side in the inequality (6), the model $M_i$ is preferred. (Gelman et al., 2004)

In the classical hypothesis test, the null hypothesis is special which requires overwhelming evidence in order to reject it. However, in the Bayesian hypothesis test framework, all hypotheses are treated equally. In this situation, assigning a large prior probability to one hypothesis or specifying a large loss from erroneously choosing the other hypothesis make it possible to simulate the sense of null hypothesis as in the frequentist hypothesis test. The left-hand side of the inequality (6) could be considered as “the test statistic” and the right-hand side of the inequality (6) could be considered as “the critical value” as in the frequentist hypothesis test framework. The critical value is estimated with particular Bayesian hypothesis tests to simulate the similar effect as the significance level.
To consider the Bayesian hypothesis test for heteroscedasticity, two different models are involved in the paper: \( M_{\text{heter}} \) is the heteroscedasticity model and \( M_{\text{hom}} \) is the model for homoscedasticity.

\[
\begin{align*}
H_0 &: \text{Homoscedasticity} \\
H_1 &: \text{Heteroscedasticity}
\end{align*}
\]

If the Bayes factor \( B_{\text{heter/hom}} = \frac{f(y^\prime | M_{\text{heter}})}{f(y^\prime | M_{\text{hom}})} > c \) (\( c \) is the critical value) is true, reject the null hypothesis.

**The heteroscedasticity Bayesian model \( M_{\text{heter}} \) and the homoscedasticity \( M_{\text{hom}} \)**

**Heteroscedasticity model \( M_{\text{heter}} \)**

The heteroscedasticity model is a hierarchical Bayesian model:

\[
\begin{align*}
y | \beta, \Sigma &\sim N(X\beta, \Sigma) \\
\beta &\sim N(\underline{\beta}, \Omega) \\
\Sigma | \alpha, w^2 &\sim \log N(X\alpha, w^2) \\
\alpha &\sim N(\underline{\alpha}, \Delta) \\
w^2 &\sim IG(\nu, \rho)
\end{align*}
\]

Where \( y \) is a vector of \( n \) dependent variables, \( X \) is an \( n \times k \) matrix of independent variables including a constant term, \( \beta \) is a vector of \( k \) parameters, \( \underline{\beta} \) is the parameters vector and \( \Omega \) is the variance and covariance matrix for the \( \beta \) prior distribution. \( \Sigma \) is the diagonal variance matrix and its prior distribution is a lognormal distribution with parameters \( \alpha \) and \( w^2 \) which have their own prior distributions. \( \underline{\alpha} \) and \( \Delta \) are the parameters of the \( \alpha \) prior distribution, while \( \nu \) and \( \rho \) are parameters of the \( w^2 \) prior distribution. \( N \) means normal distribution, \( \log N \) means lognormal distribution and IG
means inverse gamma distribution. The model is a hierarchical model because the parameters in the $\Sigma$ have their own distributions.

The data likelihood distribution:

$$f(y|\beta,\Sigma,X) = (2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}\exp[-\frac{1}{2}(y - X\beta)'\Sigma^{-1}(y - X\beta)]$$  \hspace{1cm} (8)

This is the likelihood function for linear regression. The details of every part are expressed in the model (7).

$$\Sigma = \begin{pmatrix}
\sigma_1^2 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \sigma_n^2
\end{pmatrix}$$

$\sigma_1,\sigma_2,\ldots,\sigma_n$ are different which represents the heteroscedasticity in the linear regression.

The conjugate prior for $\beta$ is

$$p(\beta) = (2\pi)^{-k/2} |\Omega|^{-1/2} \exp[-\frac{1}{2}(\beta - \beta')'\Omega^{-1}(\beta - \beta')]$$  \hspace{1cm} (9)

Where $k$ is the number of parameters including a constant.

The $\Sigma$ is assumed to have a lognormal distribution and expected value is $X\alpha$:

$$\ln\sigma^2 = X\alpha + u$$
$$u \sim N(0,w^2)$$  \hspace{1cm} (10)

Where $\sigma^2$ is vector of all the $\sigma_i^2$.

The likelihood function for $\Sigma$ is

$$f(\Sigma|\alpha,w) = (2\pi w^2)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}\exp[-\frac{1}{2w^2}(\ln\sigma^2 - X\alpha)'(\ln\sigma^2 - X\alpha)]$$

$$\ln\sigma^2 = (\ln\sigma_1^2,\ln\sigma_2^2,\ldots,\ln\sigma_n^2)$$  \hspace{1cm} (11)
The conjugate prior distribution for $\alpha$ is:

$$p(\alpha) = (2\pi)^{-\frac{k}{2}}|\Delta|^{-\frac{1}{2}} \exp[-\frac{1}{2}(\alpha - \alpha')^\top \Delta^{-1}(\alpha - \alpha)]$$  \hspace{1cm} (12)$$

The conjugate prior distribution for $w^2$ is:

$$p(w^2) = \frac{\rho^2}{\Gamma(\nu)}(w^2)^{-(\nu+1)} \exp\left(\frac{-\rho}{w^2}\right)$$  \hspace{1cm} (13)$$

By multiplying all the parts in this hierarchical model, the posterior distribution becomes

$$p(\beta, \Sigma, \alpha, w^2 | y) \propto f(y|\beta, \Sigma) \cdot p(\beta) \cdot f(\Sigma|\alpha, w^2) \cdot p(\alpha) \cdot p(\nu)$$  \hspace{1cm} (14)$$

In this situation, it is easy to get the conditional posterior distributions for all parameters and the Gibb sampling can be applied to get marginal distributions for all parameters (Gelman et al., 2004). The details of Gibb sampling are explained in the part of the Bayes factor simulation.

1) Conditioning on $\Sigma, \alpha$ and $w^2$, the conditional posterior for $\beta$ is

$$p(\beta | y, X, \Sigma, \alpha, w^2) \propto \exp[-\frac{1}{2}(y - X\beta)\Sigma^{-1}(y - X\beta)^\top \exp[-\frac{1}{2}(\beta - \beta')\Omega^{-1}(\beta - \beta')]$$  \hspace{1cm} (15)$$

This is the kernel of the normal distributions, so we have

$$\beta | y, X, \Sigma, \alpha, w^2 \sim N(\overline{\beta}, \overline{\Omega})$$
$$\overline{\Omega} = (X'\Sigma^{-1}X + \Omega^{-1})^{-1}$$
$$\overline{\beta} = \overline{\Omega}(X'\Sigma^{-1}y + \Omega^{-1}\overline{\beta})$$  \hspace{1cm} (16)$$

2) Conditioning on $\beta, \alpha$ and $w^2$, the conditional posterior for $\Sigma$ is:
\[ p(\Sigma|y,X,\beta,\alpha,w^2) \propto (2\pi)^{-n/2} |\Sigma|^{-n/2} \exp\left[-\frac{1}{2}(y-X\beta)'\Sigma^{-1}(y-X\beta)\right] \]
\[ = (2\pi)^{-n/2} |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2w^2}(\ln\sigma^2 - X\alpha)'(\ln\sigma^2 - X\alpha)\right\} \]

(17)

There is no closed form of this distribution, and the normalized constant part is impossible to calculate.

3) Conditioning on the \( \Sigma, w^2 \) and \( \beta \), the conditional posterior for \( \alpha \) is:

\[ p(\alpha|y,X,\beta,\Sigma,w^2) \propto \exp\left[-\frac{1}{2w^2}(\ln\sigma^2 - X\alpha)'(\ln\sigma^2 - X\alpha)\right] \exp\left[-\frac{1}{2}(\alpha-\alpha)'\Delta^{-1}(\alpha-\alpha)\right] \]
\[ = \exp\left\{-\frac{1}{2w^2}(\ln\sigma^2 - X\alpha)'(\ln\sigma^2 - X\alpha) + (\alpha-\alpha)'\Delta^{-1}(\alpha-\alpha)\right\} \]

(18)

This is the kernel of the normal distribution:

\[ \alpha|y,X,\Sigma,\beta,w^2 \sim N(\overline{\alpha},\overline{\Delta}) \]
\[ \overline{\Delta} = \left(\frac{1}{w^2} X'X + \Delta^{-1}\right)^{-1} \]
\[ \overline{\alpha} = \overline{\Delta}(\frac{1}{w^2} X'\ln\sigma^2 + \Delta^{-1}\alpha) \]

(19)

4) Conditioning on the \( \Sigma, \beta \) and \( \alpha \), the conditional posterior for \( w^2 \) is:

\[ p(w^2|y,X,\beta,\alpha,\Sigma) \propto (w^2)^{-\left(v+\frac{n}{2}\right)} \exp\left(-\frac{\rho}{w^2}\right) \exp\left[-\frac{1}{2w^2}(\ln\sigma^2 - X\alpha)'(\ln\sigma^2 - X\alpha)\right] \]

(20)

This is the kernel for the inverse gamma distribution:

\[ w^2|y,X,\Sigma,\beta,\alpha \sim IG(\overline{v},\overline{\rho}) \]
\[ \overline{v} = v + \frac{n}{2} \]
\[ \overline{\rho} = \overline{\rho} + \frac{1}{2}(\ln\sigma^2 - X\alpha)'(\ln\sigma^2 - X\alpha) \]

(21)
**Homoscedasticity model** $M_{\text{hom}}$

With homoscedasticity, the $\Sigma = \begin{pmatrix} \gamma^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \gamma^2 \end{pmatrix}$, the disturbances have the same variance. The likelihood function in (8) can be simplified as:

$$f(y|\beta, \gamma^2, X) = (2\pi \gamma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\gamma^2}(y - X\beta)'(y - X\beta)\right]$$

The model becomes simpler as follows:

$$y|\beta, \gamma^2 \sim N(X\beta, \gamma^2)$$
$$\beta \sim N(\beta, \Omega)$$
$$\gamma^2 \sim IG(\tau, \phi)$$

(23)

The prior for $\beta$ is the same as the one that was used in the $M_{\text{heter}}$ and $\gamma^2$ is similar to $\nu^2$ with different parameters:

$$p(\beta) = (2\pi)^{-\frac{k}{2}}|\Omega|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta - \beta)'\Omega^{-1}(\beta - \beta)\right]$$

(24)

$$p(\gamma) = \frac{\phi^{\frac{1}{2}} (\gamma^2)^{-\frac{1}{2}+1}}{\Gamma(\frac{\tau}{2})} \exp\left(-\frac{\phi}{\gamma^2}\right)$$

(25)

Conditioning on each other, the posteriors for $\beta$ and $\gamma$ are:

$$p(\beta|y, X, \gamma^2) \propto \exp\left[-\frac{1}{2\gamma^2}(y - X\beta)'(y - X\beta)\right] \exp\left[-\frac{1}{2}(\beta - \beta)'\Omega^{-1}(\beta - \beta)\right]$$

$$\beta|y, X, \gamma^2 \sim N(\overline{\beta}, \overline{\Omega})$$
$$\overline{\Omega} = \left(\frac{XX'}{\gamma^2} + \Omega^{-1}\right)^{-1}$$
$$\overline{\beta} = \overline{\Omega} \left(\frac{X'y}{\gamma^2} + \Omega^{-1}\beta\right)$$

(26)

$$p(\gamma^2|y, X, \beta) \propto (\gamma^2)^{-(\frac{\tau+1}{2})} (\gamma^2)^{-n/2} \exp\left(-\frac{\phi}{\gamma^2}\right) \exp\left[-\frac{1}{2\gamma^2}(y - X\beta)'(y - X\beta)\right]$$
\[ \gamma^2 | y, X, \beta \sim IG(\tau, \phi) \]
\[ \tau = \xi + \frac{1}{2} \]
\[ \phi = \phi + \frac{1}{2} (y - X\beta)'(y - X\beta) \]

(27)

Bayes factor simulation

In the Bayesian hypothesis test, the Bayes factor \( (6) \) plays an important role. In order to calculate the Bayes factor, the marginal likelihood is needed which is obtained by integrating over the prior distribution as in equation (11). In the high dimensional parameter space (more than three parameters in the model), it is almost impossible to do this integrating most of times.

In order to estimate the marginal likelihoods, Chib (1995) suggested a method based on the Gibbs sampler, and Chib and Jeliazkov (2001) generalized this to Metropolis-Hastings samplers. Both these two samplers are based on Markov Chain Monte Carlo simulation and the full conditional posterior is needed in order to perform the simulation. In this paper, both methods are used to estimate the marginal likelihoods.

The key insight of those methods is based on the Bayes rules.

\[
p(\theta_{M_i}|y^o,M_i) = \frac{f(y^o|M_i,\theta_{M_i})p(\theta_{M_i}|M_i)}{p(y^o|M_i)}, \text{ where } p(\theta_{M_i}|y^o,M_i) \text{ is the posterior distribution for parameters } \theta_{M_i} \text{ in the model } M_i \text{ and the meaning of other parts are described in the model expression (3).} \]

The marginal likelihood \( p(y^o|M_i) \) is actually the normalizing constant of the posterior density. Therefore, when the data is known, this is the same constant for any points inside the parameter space. The marginal likelihood is changed to the right side, and taking logarithms on both sides can avoid underflow in computation. Estimating the marginal likelihood becomes:

\[
\log[p(y^o|M_i)] = \log[f(y^o|M_i,\theta_{M_i}^*)] + \log[p(\theta_{M_i}^*|M_i)] - \log[p(\theta_{M_i}^*|y^o,M_i)]
\]

(28)

In this way, the calculation of the marginal likelihood is reduced to find the left three parts with a single point \( \theta_{M_i}^* \). This could be any point in the parameter space. Normally,
the high-density point is used to increase the efficiency. The first two parts in the left-hand side of equation (28) are easy to calculate because the distributions and the data are both known. The last part of equality (28) is mostly focused on in the simulation.

1) Simulation for marginal likelihood in Heteroscedasticity model $M_{hor}$:

$$
\log(p(y|\alpha, \beta, \Sigma, w^2, M_{hor})) = \log(f(y^o|\alpha^*, \beta^*, \Sigma^*, w^{2*})) + \log(f(\Sigma|\alpha^*, w^{2*})) \times p(\alpha^*) \times p(\beta^*) \times p(w^{2*})
$$

$$
- \log(p(\alpha^*, \beta^*, \Sigma^*, w^{2*}|y^o))
$$

(29)

In the heteroscedasticity model, the marginal likelihood for $y^o$ can be calculated by equation (29). Where * means the fixed point. If $\alpha^*, \beta^*, \Sigma^*$ and $w^{2*}$ are all fixed points, the first two parts $f(y^o|\alpha^*, \beta^*, \Sigma^*, w^{2*})$ and $f(\Sigma|\alpha^*, w^{2*}) \times p(\alpha^*) \times p(\beta^*) \times p(w^{2*})$ are just simple calculations. Chib (1995) suggests a method to estimate the posterior in the context of Gibb MCMC sampling. In the heteroscedasticity model, the posterior $p(\alpha^*, \beta^*, \Sigma^*, w^{2*}|y^o)$ needs to be estimated. Then, applying the law of total probability, we have:

$$
p(\alpha^*, \beta^*, \Sigma^*, w^{2*}|y^o) = p(\Sigma^*|y^o) \times p(\beta^*|\Sigma^*, y^o) \times p(\alpha^*|\Sigma^*, w^{2*}, y^o) \times p(w^{2*}|\Sigma^*, \alpha^*, y^o)
$$

(30)

Where $p(\Sigma^*|y^o)$ is the marginal posterior density for special point $\Sigma^*$, which can use the ordinary Gibb sampler based on the full conditional distributions for all parameters. $p(\beta^*|\Sigma^*, y^o)$ and $p(\alpha^*|\Sigma^*, y^o)$ are both the marginal posterior density for special points $\beta^*$ and $\alpha^*$ conditioning on $\Sigma^*$. $p(w^{2*}|\Sigma^*, \alpha^*, y^o)$ is the marginal posterior density for $w^{2*}$ conditioning on $\Sigma^*$ and $\alpha^*$.

The mixed stage Gibb sampler and Metropolis-Hastings (Chib and Jeliazkov, 2001) are conducted to estimate the marginal likelihood for the heteroscedasticity model.

(A) Estimate $p(\Sigma^*|y^o)$

1. Specify the initial values for $\Sigma_o, \alpha_o, \beta_o$ and $w_o^2$
2. Conduct the Metropolis-Hastings method for sampling from (17) because there is no closed form for the posterior distribution of $\Sigma$.

a) Draw the proposal points $\Sigma_p$ from an independent multivariable normal distribution $MVN(\Sigma_0, \omega)$. $\omega$ affects the acceptance rate that is the proportion of proposal $\Sigma_p$ being accepted during next steps b), c) and d).

b) Calculate

$$p = \frac{\exp\left\{-\frac{1}{2}\left[(y - X\beta_0)'\Sigma_p^{-1}(y - X\beta_0) - \frac{1}{w_0}(\ln\sigma_p^2 - X\alpha_0)'(\ln\sigma_p^2 - X\alpha_0)\right]\right\}}{\exp\left\{-\frac{1}{2}\left[(y - X\beta_0)'\Sigma_0^{-1}(y - X\beta_0) - \frac{1}{w_0}(\ln\sigma_0^2 - X\alpha_0)'(\ln\sigma_0^2 - X\alpha_0)\right]\right\}}$$

c) Draw a value $q$ from the uniform distribution $U(0,1)$, where $U$ means the uniform distribution.

d) If $p > q$, set the $\Sigma_1$ value to $\Sigma_p$. Otherwise, keep the old value $\Sigma_0$.

3. Sample $\beta_1$ from the posterior multinomial distribution (16) with the value $\Sigma_1$ from step 2.

4. Sample $\alpha_1$ from the posterior multinomial distribution (19) with the value $\Sigma_1$ from step 2 and initial value $w_0^2$.

5. Sample $w_1^2$ from the posterior inverse gamma distribution (21) with the value $\Sigma_1$ from step 2 and the value $\alpha_1$ from step 4.

6. Go back to step 2 with the values $\Sigma, \alpha_1, \beta_1$ and $w_1^2$.

7. Repeat step 2 to step 6 $N$ times including burn-in time $B$ till all the marginal distributions converge.

Burn-in time is the sufficient time to throw away before the Markov Chain is converging to the marginal distributions.

8. Find the high-density value $\Sigma^*$ in the converged marginal distribution for $\Sigma$ and its estimated density $p(\Sigma^*|y^o)$.
(B) Estimate $p(\beta^*|\Sigma^*,y^o)$

1. Sample $\beta_1$ from the posterior multi-normal distribution (16) with the special value $\Sigma^*$ from the part (A).

2. Repeat step 1 $N$ times including burn-in time $B$ till the marginal distribution converges.

3. Find the high-density value $\beta^*$ and its estimated density $p(\beta^*|\Sigma^*,y^o)$ with the converged marginal distribution.

(C) Estimate $p(\alpha^*|\Sigma^*,w^2,y^o)$

1. Sample $\alpha_1$ from the posterior multi-normal distribution (19) with the special value $\Sigma^*$ from part (A) and the initial value $w_0^2$.

2. Sample $w_1^2$ from the posterior inverse gamma distribution (21) with the value $\Sigma^*$ from part (A) and the value $\alpha_1$ from step 1.

3. Go back to step 1 with the value $w_1^2$.

4. Repeat step 1 to step 3 $N$ times including burn-in time $B$ till all the marginal distributions converge.

5. Find the high-density value $\alpha^*$ and its estimated density $p(\alpha^*|\Sigma^*,w^2,y^o)$.

(D) Estimate $p(w^2|\Sigma^*,\alpha^*,y^o)$

1. Sample $w_1^2$ from the posterior inverse gamma distribution (21) with the value $\Sigma^*$ from part (A) and the value $\alpha^*$ from part (C).

2. Repeat step 1 $N$ times including burn-in time $B$ till all the marginal distribution converges.

3. Find the high-density value $w^{2*}$ and its estimated density $p(w^{2*}|\Sigma^*,\alpha^*,y^o)$.
(E) Estimate \( p(y|\alpha, \beta, \Sigma, w^2, M_{\text{heter}}) \)

The \( p(y|\alpha, \beta, \Sigma, w^2, M_{\text{heter}}) \) is estimated by putting all the special parameters \( \Sigma^*, \alpha^*, \beta^*, w^{*2} \) and the densities \( p(\Sigma^*|y^o), p(\beta^*|\Sigma^*, y^o), p(\alpha^*|\Sigma^*, w^{*2}, y^o), p(w^{*2}|\Sigma^*, \alpha^*, y^o) \) into the equation (29) and (30).

2) Homoscedasticity model

\[
\log(p(y|\beta, \gamma^2, M_{\text{heter}})) = \log(f(y^o|\beta^*, \gamma^{*2}) + \log(p(\beta^*) \times p(\gamma^{*2}) - \log(p(\beta^*, \gamma^{*2}|y^o))
\]
\[
p(\beta^*, \gamma^{*2}|y^o) = p(\beta^*|y^o) \times p(\gamma^{*2}|\beta^*, y^o)
\]

The homoscedasticity model is simpler with only two parameters while they all have a closed form. Gibbs sampling can be used here to estimate the marginal likelihood (Gelman et al., 2004).

(A) Estimate \( p(\beta^*|y^o) \)

1. Set initial value \( \gamma_0 \).

2. Sample \( \beta_i \) from the posterior multi-normal distribution (26) with the initial value \( \gamma_0 \).

3. Sample \( \gamma_i^2 \) from the posterior inverse gamma distribution (27) with the value \( \beta_i \) from step 2.

4. Go back to step 2 with the value \( \gamma_i^2 \) from step 3.

5. Repeat step 2 to step 4 N times including burn-in time B till all the marginal distributions converge.

6. Find the high-density value \( \beta^* \) and its estimated density \( p(\beta^*|y^o) \).
(B) Estimate $p(\gamma^2|\beta^*,y^n)$

1. Sample $\gamma_1^2$ from the posterior inverse gamma distribution (27) with the value $\beta^*$ from part (A).

2. Repeat step 1 $N$ times including burn-in time $B$ till all the marginal distributions converge.

3. Find the high-density value $\gamma^2*$ and its estimated density $p(\gamma^2*|\beta^*,y^n)$.

(C) Calculate $p(y|\beta,\gamma^2,M_{\text{hom}})$ according to equation (31) in the same way as for the heteroscedasticity part.

3) Bayes factor is $B_{\text{heter/hom}} = \frac{p(y|\alpha,\beta,\Sigma,w^2,M_{\text{heter}})}{p(y|\beta,\gamma^2,M_{\text{hom}})}$, in which $p(y|\beta,\gamma^2,M_{\text{hom}})$ and $p(y|\alpha,\beta,\Sigma,w^2,M_{\text{heter}})$ are separately estimated in part 1) and part 2).

The tests for convergence

The successful Markov Chain Monte Carlo simulation is mainly based on if the simulation can actually converge to the target distributions. Before estimating the Bayes Factor, it is necessary to perform the tests for convergence and further decide the proper burn-in draws.

One way to check if the Markov Chain has converged is to see how well the chain is mixing or moving around in the parameter space. If the chain is taking a long time to move around the parameter space, then it will take longer to converge.

Some visual inspections can be conducted to see how well the chain mixing and it needs to be performed for every parameters. Trace plots, running mean plots and autocorrelation plots can be used as visual inspections. There are also a lot of diagnostics for testing convergence. The Heidelberg and Welch diagnostic (Heidelberger and Welch, 1983) together with visual diagnostic running mean and autocorrelation plot will be used in this paper.
The Heidelberg and Welch diagnostic has two parts.

The first part is a stationary test. It calculates the test statistic to accept or reject the null hypothesis: the Markov chain is from a stationary distribution. The test is applied first to the whole chain, if the null hypothesis is rejected, discarding the 10%, 20%, etc., of the chain until either the null hypothesis is accepted, or 50% of the chain is discarded which indicates a longer running is needed. If the null hypothesis is accepted, the number of iterations to keep and burn-in times are reported.

The second part is the half-width test, which uses the portion of the chain passed the stationary test to calculate a 95% confidence interval for the mean. The half-width of the interval is compared with the estimated mean. If the ratio between the half-width interval and the estimated mean is lower than a value $\xi$, the half-width test is passed. Otherwise, the length of the passed chain is not enough to estimate the sufficient accuracy mean.

The running mean plot is a plot of the iteration against the mean of all draws up to that iteration. The law of Large Numbers concerns the stability of the mean as the sample size increases. If the Chain converges, the running mean plot should show than the mean is moving tightly around a fixed value as the sample size increases.

The lag $k$ autocorrelation $\rho_k$ is the correlation between every draw and its $k$th lag.

$$
\rho_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
$$

The $k$th lag autocorrelation is expected to be smaller as $k$ increases. If autocorrelation is still relatively high for higher $k$, it indicates slow mixing.

For Metropolis-Hasting Markov Chain Monte Carlo, the acceptance rate is also important diagnostic for convergence. The acceptance rate measures the proportion of new parameters from a proposal distribution being accepted. It suggests the proper acceptance rate between 20% and 40% (Gelman et al., 2004).
The size and the power

In the frequentist hypothesis test framework

Size is the probability of rejecting $H_0$ when it is true. In order to evaluate the size of a frequentist hypothesis test, generate N times data under the null hypothesis and calculate proportion of rejections of $H_0$. Proportion should be close to the significant level $\alpha$.

Power is the probability of rejecting $H_0$ when it is false. To evaluate power, generate data under the alternative hypothesis $H_1$ N times and calculate proportion of rejections of $H_0$.

When controlling the size around the significance level, the bigger power implies better performance of a test.

In the Bayesian hypothesis test framework

Recall from the part Bayesian hypothesis test:

$H_0$: Homoscedasticity

$H_1$: Heteroscedasticity

If Bayes factor $B_{\text{heter/hom}} = \frac{f(y|M_{\text{heter}})}{f(y|M_{\text{hom}})} > c$ (c is the critical value), the null hypothesis is rejected. For simulating the significance level in the frequentist hypothesis test, the N times Bayes Factor is calculated. The critical value c corresponds to the "significance level $\alpha$" when classical test is chosen. The N should be large enough to represent a general situation. Generate Bayes Factor N times with data under null hypothesis $H_0$ and find the "critical value c". The proportion, when the Bayes Factors are bigger than c, should be equal to the significance level $\alpha$. The power is the proportion when the
Bayes Factors are larger than c with the data generated under the alternative hypothesis $H_1$ N times.

In the $M_{het}$ and $M_{hom}$ the Bayes Factor is actually affected by the sample size because the number of elements in the diagonal of $\Sigma$ in the heteroscedasticity model $M_{het}$ is growing as the sample size is growing. The bigger the size gets, the smaller the Bayes Factor will become. So the different sample sizes have different “critical value c” in our Bayesian hypothesis test.
Monte Carlo Simulation

Data generating process

In order to simplify the problem without losing generality, four independent variables from normal distributions are used in the simulation. The heteroscedasticity appeared normally in panel data in which $R^2$ is normally quite low. Therefore, $R^2$ is chosen around $30\%$ to control the coefficients $\beta$ and $\alpha$ in (32), where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ and $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

Here is the model used to generate data:

$$
\begin{align*}
    y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i \\
    \varepsilon_i &\sim N(0, \sigma_i) \\
    \sigma_i^2 &= \exp(\alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4})
\end{align*}
$$

To make things simple, the parameter $\beta = (1,1,1,1,1)$ is chosen here and the $x_1 \sim N(1,1)$, $x_2 \sim N(1,1.2)$, $x_3 \sim N(1,1.5)$ and $x_4 \sim N(1,1.8)$ are generated, where N stands for normal distribution. In order to control the $R^2$ around $30\%$, the variance of the disturbances should be around $18$. In other words, the expected value of the variance in model (32) $\exp(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)$ is around $18$. So $\alpha = (2,0.3,0.2,0.25,0.25)$ is chosen in the simulation study for the data with heteroscedasticity. With this parameter $\alpha$, even when no heteroscedasticity is in the data, the $R^2$ is still quite low around $50\%$.

Three different comparisons are performed in our study: different sample size, normality assumption and the structure of variances.

1) Small sample size behavior is normally interesting because the asymptotically distribution of the test statistic is violated when the sample size is small. In this special case, the degrees of freedom can grow dramatically if the test statistic is estimated by regressing variances against all the independent variables, their squared forms and their cross products.
a) Generate $x_1, x_2, x_3$ and $x_4$ with different sample size 20, 50, 100.

b) The $x_1, x_2, x_3$ and $x_4$ from step a) together with the parameter
   \[ \alpha = (2, 0.3, 0.2, 0.25, 0.25) \]
   create variances for heteroscedastic data and
   \[ \alpha = (2, 0, 0, 0, 0) \]
   for homoscedastic data according to (32).

c) The disturbances $\varepsilon_i \sim N(0, \sigma_i^2)$ are generated with variances $\sigma_i^2$ from step b).

d) $\beta = (1,1,1,1)$, the $x_1, x_2, x_3$ and $x_4$ from step a), and the disturbances from step c)
   finally generate the $y$ for heteroscedastic data and for homoscedastic data.

2) The normality assumption is important in some frequentist hypothesis tests. In this part, the new model violates this assumption. Sample size 50 and 100 is chosen to compare between the two models. The new model is:

\[
\begin{align*}
  y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \\
  \varepsilon_i &\sim t(p) \\
  \sigma_i^2 &= \exp(\alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2})
\end{align*}
\]

(33)

Instead of disturbances from a normal distribution, the $t$ distribution with fixed $p$ degree of freedom is used to generate disturbances. In order to get the special variance $\sigma_i^2$, the standard $t$ distribution needs to be scaled with the factor $\sigma_i \sqrt{(p-2)/p}$.

a) With sample size 50 and 100, $x_1, x_2, x_3$ and $x_4$ are generated in the same manner as before.

b) The $x_1, x_2, x_3$ and $x_4$ from a) together with the parameter
   \[ \alpha = (2, 0.3, 0.2, 0.25, 0.25) \]
   create variances for heteroscedastic data and
   \[ \alpha = (2, 0, 0, 0, 0) \]
   for homoscedastic data according to (33).
c) Generate the disturbances from the scaled t distribution:

\[ \varepsilon_i \sim rt(1, df = p) \times \sigma_i \times \sqrt{\frac{(p - 2)}{p}} \],

rt is the random sample from the t distribution and p = 9 is chosen in this comparison.

d) \( \beta = (1,1,1,1,1) \), the \( x_1, x_2, x_3 \) and \( x_4 \) from step a), and disturbances from step c) finally generate the dependent variable y for heteroscedastic data and for homoscedastic data.

3) The White test can only test whether heteroscedasticity exists but have no clue which independent variables create the problem. It is interesting to know that if the right guessing of the variances’ structure in the Breusch-Pagan Test and Koenker-Basset Test can give bigger power.

In this part, the data is generated in the same way as the first comparison but with only sample size 50 and 100. In the step b), the \( \alpha = (2,0.3,0,0,0.25) \) is used to generate variance for heteroscedastic data.

The important part in this comparison is that the White test still tests on all the independent variables and their possible cross products and squares. But the Breusch-Pagan Test and Koenker-Basset Test choose the independent variable \( x_1 \) and \( x_4 \) which is the reason for heteroscedasticity as we know from the data generated process, \( x_1^2, x_4^2 \) and \( x_1x_4 \) is also included in the test statistic.

Parameters for the prior distributions in Bayesian models

Before the simulation for estimating the Bayes factor or further the hypothesis test, the prior parameters should be chosen according to the pre-researches, the common knowledge, and other information that helps to guess the prior parameters. In this paper the data with heteroscedasticity are generated according to the simulation part with sample size 100.
Here is information for the data randomly generated according to the data generating process:

<table>
<thead>
<tr>
<th></th>
<th>$y$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td>33.77</td>
<td>0.79</td>
<td>1.33</td>
<td>1.90</td>
<td>3.49</td>
</tr>
<tr>
<td>Expectation</td>
<td>4.90</td>
<td>1.13</td>
<td>0.99</td>
<td>0.80</td>
<td>0.69</td>
</tr>
</tbody>
</table>

a) The parameters $\beta$ and $\Omega$ for the prior $\beta$ distribution.

The Data generated process is according to model (1), so:

$$\text{Expect}(y) = \beta_0 + \beta_1 \text{Expect}(x_1) + \beta_2 \text{Expect}(x_2) + \beta_3 \text{Expect}(x_3) + \beta_4 \text{Expect}(x_4)$$
$$\text{Var}(y) = \beta_1^2 \text{Var}(x_1) + \beta_2^2 \text{Var}(x_2) + \beta_3^2 \text{Var}(x_3) + \beta_4^2 \text{Var}(x_4) + \text{Var}(\varepsilon)$$

Where Var() means variance and Expect() is expectation.

There is no special information about the data except that the data are simulated according to the panel data, so the variances of $x$ together is normally between 10% and 60% of variance $y$. 30% will be chosen and that means the variance together for $x$ is around 10.13. If the coefficients are assumed to be equal, $\beta_1, \beta_2, \beta_3$ and $\beta_4$ are 1.16. The $\beta_0 = 0.71$ is given by subtracting $\beta_1 \exp(x_1) + \beta_2 \exp(x_2) + \beta_3 \exp(x_3) + \beta_4 \exp(x_4)$ from $\exp(y)$. The guessing is rough without more information, so the large variance is applied to prior parameters, which can make them rather uninformative. In other word, the data dominate the posterior distribution. The least squares estimates of the variance are borrowed here and the chosen prior $\sigma_\beta^2 = 20 \times \text{diag}(X'X)^{-1} \text{var}(\varepsilon)$ which is around $\sigma_{\beta_0}^2 = 5.25, \sigma_{\beta_1}^2 = 1.83, \sigma_{\beta_2}^2 = 1.08, \sigma_{\beta_3}^2 = 0.75$ and $\sigma_{\beta_4}^2 = 0.41$. 20 times is for making the variances big and taking the diagonal of the matrix in order to just get the variances without the covariance $\Omega = \text{diag}(\sigma_\beta)$. The same parameters for the $\beta$ prior distribution are used in both the heteroscedasticity and the homoscedasticity model. $\text{diag()}$ is the express in R which takes the diagonal of the square matrix to a vector, or makes a vector to a square matrix.
2) The parameter $\tau$ and $\varphi$ for the $\gamma^2$ prior distribution in the homoscedasticity model

According to the posterior distribution of $\gamma^2$ (27):

$$\tilde{\tau} = \tau + \frac{\varphi}{\tau}$$
$$\tilde{\varphi} = \varphi + \frac{1}{2}(y - X\beta)'(y - X\beta)$$

The prior distribution (25) give the information of our guessing of expectation and variance of the $\gamma^2$:

$$\exp(\gamma^2) = \varphi / (\tau - 1)$$
$$\text{var}(\gamma^2) = \varphi / (\tau - 1)^2(\tau - 2)$$
$$\tau > 2$$

The prior $\tau$ and $\varphi$ will give less information compared with the data information. $\tau = 2.1$ is used to give less than 20% information when the sample size is 20 and less when the sample size becomes larger. The expectation of the $\gamma^2$ is 23.64 which is around 70% of the total variance. $\varphi = 23.64 \times (\tau - 1) \approx 26.00$ gives less information in the posterior distribution and bigger variance in the prior distribution. In this way, the data dominate the information instead of the prior distributions.

3) The parameters $\alpha$ and $\Delta$ for the $\alpha$ prior distribution in the heteroscedasticity model

When heteroscedasticity is assumed, each disturbance has different variance. Because there is only one single observation for every disturbance, the square of every disturbance is used to guess the variance for itself. Letting $\sigma^2 = \varepsilon^2$, the new linear regression model can be used:

$$\varepsilon = y - X\beta$$
$$\ln\sigma^2 = \ln\varepsilon^2 = X\alpha + u, \ u \sim N(0, w^2)$$

$$\text{Expect}(\ln\varepsilon^2) = \alpha_0 + \alpha_1\text{Expect}(x_1) + \alpha_2\text{Expect}(x_2) + \alpha_3\text{Expect}(x_3) + \alpha_4\text{Expect}(x_4)$$
$\text{Expect}(\ln \varepsilon^2)$ is around 3.10 by calculating from $y, X$ and $\beta$, where $\text{Expect}()$ means expectation.

$\alpha_0 + 1.13\alpha_1 + 0.99\alpha_2 + 0.80\alpha_3 + 0.69\alpha_4 \approx 3.10$

$\alpha_0$ should be chosen in this way even when the other parameters $\alpha_1, \alpha_2, \alpha_3$ and $\alpha_4$ are zero, the variance of the disturbance can still take quite a big part of the variance of $y$ as in the panel data. $\alpha_0 = 2.10$ is chosen here. Assuming other $\alpha$ except $\alpha_0$ are equal, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 \approx 0.28$. The variances of $\alpha$ use the same way to estimate as the variance of $\beta, \sigma^2_\varepsilon = 20 \times \text{diag}(XX)^{-1} \text{var}(\ln \varepsilon^2 - X\alpha)$. The prior guessing is $\sigma^2_\alpha = (4.06, 1.42, 0.84, 0.58, 0.32)$. 20 times the estimate variance makes the prior less informative.

4) The parameters $\nu$ and $\rho$ for the $w^2$ prior distribution

The parameters from posterior distribution (21) are:

$\bar{\nu} = \nu + \frac{s}{2}$

$\bar{\rho} = \rho + \frac{1}{2}(\ln \sigma^2 - X\alpha)'(\ln \sigma^2 - X\alpha)$

The expectation and variance from the prior distribution (13):

$\text{exp}(w^2) = \frac{\rho}{(\nu - 1)}$

$\text{var}(w^2) = \frac{\rho}{(\nu - 1)^2(\nu - 2)}$

$\nu > 2$

$\nu = 2.1$ is chosen in order to supply less information to the posterior distribution comparing with the data. This is actually zero in the data generating part. So the expectation of $w^2$ should be quite small. If $\text{exp}(w^2) = 0.1$, then the $\rho$ is 0.11.
Starting points in Bayesian Markov Chain Monte Carlo simulation

Starting points are chosen close to the real data generated parameters in order to shorten the convergence process.

\[
\alpha = (1.5, 0.3, 0.2, 0.25, 0.25) \\
\beta = (1, 1, 1, 1) \\
w^2 = 0.001 \\
\sigma_i^2 = \exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \alpha_4 x_{4i})
\]

For the meanings of the parameters see the section: the heteroscedasticity Bayesian model \( M_{\text{heter}} \) and the homoscedasticity \( M_{\text{hom}} \).

Other parameters in simulation study

The significance level 0.05 is used in all the tests and the number of replicates is 1000 for calculating the size and the power of tests.

Markov Chain Monte Carlo simulation for Bayes Factor runs 5000 times. The Acceptance rate in the Metropolis-Hasting is controlled around 30% in the simulation study.
Monte Carlo Results

From the visual inspections of the Running mean plot and the Heidelberg and Welch diagnostic, all the parameters from Heteroscedasticity and Homoscedasticity Bayesian model are converging after some draws. The Autocorrelation plot indicates the good mixing. In this study, 2000 time's burn-in draws is chosen out of total running time 5000. The results from the Heidelberg and Welch diagnostic, running mean plots and autocorrelation plots are listed in Appendix.

For different sample sizes, the Bayesian hypothesis test has different “critical value $c$": $c_{20} = -47.47, c_{50} = -147.52$ and $c_{100} = -309.80$. Those critical values are in the logarithm form. The results come from the simulation.

### Table 1  Power and size for different sample sizes

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
<td>BP</td>
</tr>
<tr>
<td>20</td>
<td>0.044</td>
<td>0.045</td>
</tr>
<tr>
<td>50</td>
<td>0.039</td>
<td>0.059</td>
</tr>
<tr>
<td>100</td>
<td>0.054</td>
<td>0.065</td>
</tr>
</tbody>
</table>

W: the white test, BP: Breusch-Pagan test, KB: Koenker-Basset test, BY: Bayesian test

In this case, the Bayesian hypothesis test is more complex and less powerful compared with the frequentist hypothesis tests. The number of the parameters in the heteroscedasticity model is much more than in the homoscedasticity model, which makes the heteroscedasticity model less likely to be chosen even if it may be the “correct” model. In the same time, the number of parameters in the heteroscedasticity model grows with the sample size because of $\Sigma$.

When controlling the size around 0.05, the powers in all tests increase as the sample size becomes large. Among all the tests above, the White test and the Breusch-Pagan test
behave best. We expect the Koenker-Basset test performs better when the sample size is even larger. The Koenker-Basset test fails to detect heteroscedasticity when the sample size is as small as 20. The failure can be caused by the fact that the degrees of freedom are too big compared with the sample size.

### Table 2  Normality assumption

<table>
<thead>
<tr>
<th>Test/Sample</th>
<th>Size</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>T</td>
<td>N</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>W / 50</td>
<td>0.039</td>
<td>0.058</td>
<td>0.453</td>
<td>0.366</td>
<td></td>
</tr>
<tr>
<td>W / 100</td>
<td>0.054</td>
<td>0.042</td>
<td>0.877</td>
<td>0.746</td>
<td></td>
</tr>
<tr>
<td>BP / 50</td>
<td>0.059</td>
<td>0.162</td>
<td>0.449</td>
<td>0.505</td>
<td></td>
</tr>
<tr>
<td>BP / 100</td>
<td>0.065</td>
<td>0.227</td>
<td>0.803</td>
<td>0.819</td>
<td></td>
</tr>
<tr>
<td>KB / 50</td>
<td>0.038</td>
<td>0.072</td>
<td>0.263</td>
<td>0.239</td>
<td></td>
</tr>
<tr>
<td>KB / 100</td>
<td>0.058</td>
<td>0.080</td>
<td>0.624</td>
<td>0.480</td>
<td></td>
</tr>
<tr>
<td>BY / 50</td>
<td>0.050</td>
<td>0.072</td>
<td>0.188</td>
<td>0.184</td>
<td></td>
</tr>
<tr>
<td>BY / 100</td>
<td>0.050</td>
<td>0.042</td>
<td>0.216</td>
<td>0.230</td>
<td></td>
</tr>
</tbody>
</table>

W: the white test, BP: Breusch-Pagan test, KB: Koenker-Basset test, BY: Bayesian test
N: Normal distributed disturbance T: T distributed disturbance

Changing the distribution of disturbances from normal to t distribution does not affect the size and the power of the Bayesian hypothesis test and the White test. By contrast, this change affects the size of Breusch-Pagan test badly as we suspected, the Breusch-Pagan test is more sensitive to the normality assumption. Table 2 also shows that the Koenker-Basset test has less power when the disturbances are t-distributed.
Table 3  Structure of variances and test variables

<table>
<thead>
<tr>
<th>Test</th>
<th>Sample size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DG $\alpha = (2,0.3,0.2,0.25,0.25)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DG $\alpha = (2,0.3,0,0,0.25)$</td>
<td></td>
</tr>
<tr>
<td>Test variable in BP, KB: $x_1, x_4$</td>
<td>Test variable in BP, KB: $x_1, x_4$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.486</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.886</td>
</tr>
<tr>
<td>W</td>
<td>50</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.722</td>
</tr>
<tr>
<td>BP</td>
<td>50</td>
<td>0.270</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.528</td>
</tr>
<tr>
<td>KB</td>
<td>50</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.216</td>
</tr>
<tr>
<td>BY</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

W: the white test, BP: Breusch-Pagan test, KB: Koenker-Basset test, BY: Bayesian test
DG: data generate process

The Breusch-Pagan and the Koenker-Basset tests both request reasonably guessing the structure of variances. The third column shows that when the guessing is wrong, the power of these two tests is much smaller comparing with the White test that includes all the independent variables, their squared form and their cross products form. However, the fourth column proves that when the guessing is right, these two tests are more powerful than the White test.
Conclusion and Discussion

We have presented a Bayesian hypothesis and three often used frequentist hypothesis tests: the White test, the Breusch-Pagan test and the Koenker-Basset test.

Frequentist hypothesis tests are effective and easy to apply. Comparisons between the frequentist hypothesis tests show that their performances vary in different situations. Choosing the proper test in the proper situation improves the power of the test.

The Bayesian hypothesis test seems unpractical to apply and behaves poorly in this case. However, the Bayesian hypothesis test may be improved by adding more informative prior or modified by eliminating the number of parameters.

The Bayesian heteroscedasticity model in our paper can be possibly applied in other situations such as estimating the parameters or the structure of variances. It leaves questions for future study: is it possible to decide which model is better under the condition that one of the models has much more parameters than another? And how?
References


Appendix

A1  autocorrelation plot for $\beta_{het}$ (20 sample size)

A2  autocorrelation plot for $\beta_{nom}$ (20 sample size)
A3  autocorrelation plot for $\alpha_{\text{het}}$ (20 sample size)

A4  autocorrelation plot for $w_{\text{het}}^2$ (20 sample size)

A5  autocorrelation plot for $\gamma_{\text{het}}^2$ (20 sample size)
A6  autocorrelation plot for $\beta_{\text{heter}}$ (50 sample size)

A7  autocorrelation plot for $\beta_{\text{hom}}$ (50 sample size)
A8  autocorrelation plot for $\alpha_{heter}$ (50 sample size)

A9  autocorrelation plot for $w_{heter}^2$ (50 sample size)

A10 autocorrelation plot for $\gamma_{heter}^2$ (50 sample size)
A11  autocorrelation plot for $\beta_{\text{heter}}$ (100 sample size)

A12  autocorrelation plot for $\beta_{\text{hom}}$ (100 sample size)
A13  autocorrelation plot for $\alpha_{\text{heter}}$ (100 sample size)

A14  autocorrelation plot for $w_{\text{heter}}^2$ (100 sample size)

A15  autocorrelation plot for $\gamma_{\text{heter}}^2$ (100 sample size)
B1  Running Mean plot for $\beta_{het}$ (20 sample size)

B2  Running Mean plot for $\beta_{hom}$ (20 sample size)
B3 Running Mean plot for $\alpha_{\text{heter}}$ (20 sample size)

B4 Running Mean plot for $w^2_{\text{heter}}$ and $\gamma^2_{\text{hom}}$ (20 sample size)
B5  Running Mean plot for $\beta_{\text{het}}$ (50 sample size)

B6  Running Mean plot for $\beta_{\text{hom}}$ (50 sample size)
B7  Running Mean plot for $\alpha_{heter}$ (50 sample size)

B8  Running Mean plot for $w_{heter}^2$ and $\gamma_{hom}^2$ (50 sample size)
B9  Running Mean plot for $\beta_{\text{heter}}$ (100 sample size)

B10  Running Mean plot for $\beta_{\text{homo}}$ (100 sample size)
B11  Running Mean plot for $\alpha_{heter}$ (100 sample size)

B12  Running Mean plot for $w_{heter}^2$ and $\gamma_{hom}^2$ (100 sample size)
### C1 Heidelberg and Welch diagnostic for $\beta_{het}$ (20 sample size)

<table>
<thead>
<tr>
<th>Stationarity start</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>test</td>
<td>iteration</td>
</tr>
<tr>
<td>var1 passed</td>
<td>1</td>
</tr>
<tr>
<td>var2 passed</td>
<td>1</td>
</tr>
<tr>
<td>var3 passed</td>
<td>1</td>
</tr>
<tr>
<td>var4 passed</td>
<td>1</td>
</tr>
<tr>
<td>var5 passed</td>
<td>502</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfwidth Mean Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1 passed</td>
</tr>
<tr>
<td>var2 passed</td>
</tr>
<tr>
<td>var3 passed</td>
</tr>
<tr>
<td>var4 passed</td>
</tr>
<tr>
<td>var5 passed</td>
</tr>
</tbody>
</table>

### C2 Heidelberg and Welch diagnostic for $\beta_{hom}$ (20 sample size)

<table>
<thead>
<tr>
<th>Stationarity start</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>test</td>
<td>iteration</td>
</tr>
<tr>
<td>var1 passed</td>
<td>1</td>
</tr>
<tr>
<td>var2 passed</td>
<td>1</td>
</tr>
<tr>
<td>var3 passed</td>
<td>1</td>
</tr>
<tr>
<td>var4 passed</td>
<td>1</td>
</tr>
<tr>
<td>var5 passed</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfwidth Mean Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1 passed</td>
</tr>
<tr>
<td>var2 passed</td>
</tr>
<tr>
<td>var3 passed</td>
</tr>
<tr>
<td>var4 passed</td>
</tr>
<tr>
<td>var5 passed</td>
</tr>
</tbody>
</table>

### C3 Heidelberg and Welch diagnostic for $\alpha_{het}$ (20 sample size)

<table>
<thead>
<tr>
<th>Stationarity start</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>test</td>
<td>iteration</td>
</tr>
<tr>
<td>var1 passed</td>
<td>1</td>
</tr>
<tr>
<td>var2 passed</td>
<td>1</td>
</tr>
<tr>
<td>var3 passed</td>
<td>1</td>
</tr>
<tr>
<td>var4 passed</td>
<td>1</td>
</tr>
<tr>
<td>var5 passed</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfwidth Mean Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1 passed</td>
</tr>
<tr>
<td>var2 passed</td>
</tr>
<tr>
<td>var3 passed</td>
</tr>
<tr>
<td>var4 passed</td>
</tr>
<tr>
<td>var5 passed</td>
</tr>
</tbody>
</table>
C4  Heidelberg and Welch diagnostic for $W_{het}^2$ and $\gamma_{het}^2$ (20 sample size)

<table>
<thead>
<tr>
<th>Test</th>
<th>Iteration</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>testw20</td>
<td>502</td>
<td>0.0745</td>
</tr>
<tr>
<td>testvariance20</td>
<td>1</td>
<td>0.3685</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>Halfwidth Mean</th>
<th>Halfwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>testw20</td>
<td>0.0143</td>
<td>0.0002</td>
</tr>
<tr>
<td>testvariance20</td>
<td>38.2045</td>
<td>0.3644</td>
</tr>
</tbody>
</table>

C5  Heidelberg and Welch diagnostic for $\beta_{het}$ (50 sample size)

<table>
<thead>
<tr>
<th>Test</th>
<th>Iteration</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1</td>
<td>1</td>
<td>0.816</td>
</tr>
<tr>
<td>var2</td>
<td>1</td>
<td>0.390</td>
</tr>
<tr>
<td>var3</td>
<td>1</td>
<td>0.305</td>
</tr>
<tr>
<td>var4</td>
<td>1</td>
<td>0.956</td>
</tr>
<tr>
<td>var5</td>
<td>1</td>
<td>0.428</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>Halfwidth Mean</th>
<th>Halfwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1</td>
<td>1.423</td>
<td>0.01575</td>
</tr>
<tr>
<td>var2</td>
<td>0.997</td>
<td>0.00999</td>
</tr>
<tr>
<td>var3</td>
<td>1.047</td>
<td>0.00998</td>
</tr>
<tr>
<td>var4</td>
<td>0.826</td>
<td>0.00847</td>
</tr>
<tr>
<td>var5</td>
<td>1.167</td>
<td>0.00562</td>
</tr>
</tbody>
</table>

C6  Heidelberg and Welch diagnostic for $\beta_{hom}$ (50 sample size)

<table>
<thead>
<tr>
<th>Test</th>
<th>Iteration</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1</td>
<td>1</td>
<td>0.861</td>
</tr>
<tr>
<td>var2</td>
<td>1</td>
<td>0.942</td>
</tr>
<tr>
<td>var3</td>
<td>1</td>
<td>0.692</td>
</tr>
<tr>
<td>var4</td>
<td>1</td>
<td>0.387</td>
</tr>
<tr>
<td>var5</td>
<td>1</td>
<td>0.946</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>Halfwidth Mean</th>
<th>Halfwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1</td>
<td>1.638</td>
<td>0.02753</td>
</tr>
<tr>
<td>var2</td>
<td>0.651</td>
<td>0.01439</td>
</tr>
<tr>
<td>var3</td>
<td>1.027</td>
<td>0.01240</td>
</tr>
<tr>
<td>var4</td>
<td>0.677</td>
<td>0.00965</td>
</tr>
<tr>
<td>var5</td>
<td>1.399</td>
<td>0.00916</td>
</tr>
</tbody>
</table>
### C7 Heidelberg and Welch diagnostic for $\alpha_{het}$ (50 sample size)

<table>
<thead>
<tr>
<th>Stationarity start</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>test</td>
<td>iteration</td>
</tr>
<tr>
<td>vor1 passed</td>
<td>1</td>
</tr>
<tr>
<td>vor2 passed</td>
<td>1</td>
</tr>
<tr>
<td>vor3 passed</td>
<td>1</td>
</tr>
<tr>
<td>vor4 passed</td>
<td>1</td>
</tr>
<tr>
<td>vor5 passed</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfwidth Mean Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>vor1 passed</td>
</tr>
<tr>
<td>vor2 passed</td>
</tr>
<tr>
<td>vor3 passed</td>
</tr>
<tr>
<td>vor4 passed</td>
</tr>
<tr>
<td>vor5 passed</td>
</tr>
</tbody>
</table>

### C8 Heidelberg and Welch diagnostic for $w_{het}^2$ and $\gamma_{hoen}^2$ (50 sample size)

<table>
<thead>
<tr>
<th>Stationarity start</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>test</td>
<td>iteration</td>
</tr>
<tr>
<td>testw50 passed</td>
<td>1</td>
</tr>
<tr>
<td>testvariance50 passed</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfwidth Mean Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>testw50 passed</td>
</tr>
<tr>
<td>testvariance50 passed</td>
</tr>
</tbody>
</table>

### C9 Heidelberg and Welch diagnostic for $\beta_{het}$ (100 sample size)

<table>
<thead>
<tr>
<th>Stationarity start</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>test</td>
<td>iteration</td>
</tr>
<tr>
<td>vor1 passed</td>
<td>1</td>
</tr>
<tr>
<td>vor2 passed</td>
<td>1</td>
</tr>
<tr>
<td>vor3 passed</td>
<td>1</td>
</tr>
<tr>
<td>vor4 passed</td>
<td>1</td>
</tr>
<tr>
<td>vor5 passed</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfwidth Mean Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>vor1 passed</td>
</tr>
<tr>
<td>vor2 passed</td>
</tr>
<tr>
<td>vor3 passed</td>
</tr>
<tr>
<td>vor4 passed</td>
</tr>
<tr>
<td>vor5 passed</td>
</tr>
</tbody>
</table>
### C10  Heidelberg and Welch diagnostic for $\beta_{\text{hom}}$ (100 sample size)

<table>
<thead>
<tr>
<th>Stationarity start test</th>
<th>p-value</th>
<th>iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1 passed</td>
<td>0.141</td>
<td>1</td>
</tr>
<tr>
<td>var2 passed</td>
<td>0.727</td>
<td>1</td>
</tr>
<tr>
<td>var3 passed</td>
<td>0.324</td>
<td>1</td>
</tr>
<tr>
<td>var4 passed</td>
<td>0.456</td>
<td>1</td>
</tr>
<tr>
<td>var5 passed</td>
<td>0.755</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfwidth Mean</th>
<th>Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1 passed</td>
<td>0.472</td>
</tr>
<tr>
<td>var2 passed</td>
<td>1.129</td>
</tr>
<tr>
<td>var3 passed</td>
<td>1.922</td>
</tr>
<tr>
<td>var4 passed</td>
<td>0.816</td>
</tr>
<tr>
<td>var5 passed</td>
<td>1.444</td>
</tr>
</tbody>
</table>

### C11  Heidelberg and Welch diagnostic for $\alpha_{\text{heter}}$ (100 sample size)

<table>
<thead>
<tr>
<th>Stationarity start test</th>
<th>p-value</th>
<th>iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1 passed</td>
<td>0.907</td>
<td>1</td>
</tr>
<tr>
<td>var2 passed</td>
<td>0.431</td>
<td>1</td>
</tr>
<tr>
<td>var3 passed</td>
<td>0.185</td>
<td>1</td>
</tr>
<tr>
<td>var4 passed</td>
<td>0.709</td>
<td>1</td>
</tr>
<tr>
<td>var5 passed</td>
<td>0.403</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfwidth Mean</th>
<th>Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>var1 passed</td>
<td>1.50</td>
</tr>
<tr>
<td>var2 passed</td>
<td>0.30</td>
</tr>
<tr>
<td>var3 passed</td>
<td>0.20</td>
</tr>
<tr>
<td>var4 passed</td>
<td>0.25</td>
</tr>
<tr>
<td>var5 passed</td>
<td>0.25</td>
</tr>
</tbody>
</table>

### C12  Heidelberg and Welch diagnostic for $\chi^2_{\text{heter}}$ and $\chi^2_{\text{hom}}$ (100 sample size)

<table>
<thead>
<tr>
<th>Stationarity start test</th>
<th>p-value</th>
<th>iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>testwv100 passed</td>
<td>0.286</td>
<td>502</td>
</tr>
<tr>
<td>testvariance100 passed</td>
<td>0.730</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfwidth Mean</th>
<th>Halfwidth test</th>
</tr>
</thead>
<tbody>
<tr>
<td>testwv100 passed</td>
<td>0.00233</td>
</tr>
<tr>
<td>testvariance100 passed</td>
<td>25.91182</td>
</tr>
</tbody>
</table>