International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 3, May-June-2024, ISSN (Online): 2395-566X

Practical Application of the Sup-Wald Test in Regression Models Using R

Research Scholar Siddamsetty Upendra, Dr. R. Abbaiah, Dr. P. Balasiddamuni, Dr. K. Murali
Department of Statistics, S V University, Tirupathi, India

Abstract- The Sup-Wald test is an important tool in regression analysis, especially for evaluating the joint significance of coefficients in a model. This paper presents a practical and descriptive study of the Sup-Wald test, focusing on its application to the simple linear regression model. We define hypotheses, elucidate the calculation of the Wald statistic, examine its distribution, determine critical values, and interpret results. Emphasis is placed on Ordinary Least Squares (OLS) estimation and variance precision. A practical illustration is presented through an R program, offering researchers hands-on guidance for implementing the Sup-Wald test in real-world scenarios. This paper provides practitioners with a clear understanding and practical skills for utilizing the Sup-Wald test in regression analysis.

Index Terms- Regression Analysis, Model Specification, Wald Statistic, Joint Significance, OLS Estimation, Simple Linear Regression, Hypothesis Testing, R Programming.

I. INTRODUCTION

Regression analysis serves as a fundamental tool in extracting valuable insights from data, allowing researchers to understand relationships between variables. The Sup-Wald test, an essential component in regression analysis, plays a pivotal role in assessing the joint significance of coefficients within a model.

This statistical test, anchored by the Wald statistic, provides a robust method for model specification. While the theoretical underpinnings of the Sup-Wald test are well-established, this paper endeavors to bridge the gap between theory and practice by offering a practical exploration of its implementation using the R programming language.

Objectives

Theoretical Understanding
Investigate the foundational principles of the Sup-Wald test, elucidating its theoretical underpinnings and significance in regression analysis.

Practical Application
Demonstrate the application of the Sup-Wald test through a simple linear regression model, focusing on assessing the joint significance of a single coefficient.

Implementation in R
Provide a step-by-step guide for implementing the Sup-Wald test in R, emphasizing hands-on practicality for researchers and analysts.

Interpretation and Insights
Equip readers with the skills to interpret Sup-Wald test results, enabling a deeper understanding of joint significance and its implications for regression models.

Real-World Relevance
Illustrate the practical relevance of the Sup-Wald test through a structured R program, showcasing its applicability in real-world scenarios.

By achieving these objectives, this paper aims to empower researchers with both theoretical insights and practical skills, encouragement a comprehensive understanding of the Sup-Wald test's role in regression analysis.

Let's consider a simple linear regression model with one dependent variable (Y) and one independent variable (X). The Sup-Wald test for this simple model will assess the joint significance of the coefficient β_1 .

Simple Linear Regression Model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Where:
Yi is the dependent variable for ith observation.
X1i is the included independent variable.
β_0 and β_1 are the coefficients to be estimated.
ϵ_i is the error term.

The OLS method is used to estimate the coefficients

International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 3, May-June-2024, ISSN (Online): 2395-566X

β_0 and β_1 in linear regression model. OLS estimates are obtained by minimizing the sum of the squared differences between the observed values of Y and the values predicted by the regression equation.

The OLS estimation of □0

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$$

and

$\hat{\beta}_1$ = estimate of the slope.
$\hat{\beta}_0$ = estimate of the intercept.
n = number of data points.

Here, Xi and Yi are the data points. X and Y are the sample means of the independent variable and dependent variable, respectively.

II. FITTED MODEL

Based on OLS estimates, the fitted model represents the estimated relationship between the dependent and independent variables. This is specified by a simple linear regression with estimated coefficients:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

This formula represents the slope coefficient that minimizes the sum of squared differences between the observed values of the dependent variable (Y) and the values predicted by the linear regression model.

Variance of (β_1^):

The variance of the estimator β_1 ˆ is given by the formula:

$$Var(\beta_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Where:

σ^2 is the variance of the error term □,
n is the number of observations,
Xi is the value of the independent variable for observation i,
X is the mean of the independent variable.

The formula for the variance of β_1 ˆ is derived using the properties of the OLS estimator. The OLS estimate of β_1 is obtained by minimizing the sum of squared differences between the observed and predicted values of the dependent variable. The variance formula involves the error term $\in_I$ and the sum of squared differences in the independent variable.

The formula reflects the precision of the estimate more precise estimate. β_1 A smaller variance indicates a more precise estimate

It's important to note that in practice, σ^2 is unknown and needs to be estimated. The estimated variance of β_1 is often replaced with its unbiased estimator, which involves substituting the residual sum of squares (RSS) and the degrees of freedom into the formula. The unbiased estimator is given by

$$Var(\hat{\beta}_1) = \frac{RSS}{(n-2)\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Where:
RSS is the residual sum of squares
This unbiased estimator of the variance is commonly used in practice for inference on the slope coefficient.

The Residual Sum of Squares (RSS) is a measure of the difference between the observed values and the values predicted by the regression model. In the context of a simple linear regression model, the RSS is defined as the sum of the squared differences between the observed dependent variable values (Yi) and the predicted values (Yˆ i)

$$RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Where:
n is the number of observations,
Yi is the observed value of the dependent variable for observation \(i\),
Yˆ I is the predicted value of the dependent variable for observation \(i\) from the regression model.
The predicted values Yˆ I are obtained by substituting the values of the independent variable (Xi) into the estimated

coefficients from the regression model. In the case of a simple linear regression model:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Now, substitute this expression for Y^ i into the RSS formula:

$$RSS = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2$$

The goal of the ordinary least squares (OLS) method is to minimize the RSS by choosing the values of the coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the sum of squared differences between the observed and predicted values.

The estimated coefficients^ and $\hat{\beta}_0$ and β_1 are obtained by minimizing the RSS. Once these estimates are obtained, you can use them to calculate the predicted values Y^ i and then compute the RSS.

Sup-Wald Test Hypotheses: Null Hypothesis (H0):
H_0: $\beta_1 = 0$ (The slope coefficient is jointly equal to zero).
Alternative Hypothesis (H1):
H_1: $\beta_1 \neq 0$ (The slope coefficient is jointly different from zero).

Sup-Wald Test Statistic:
The Wald statistic for the Sup-Wald test in this case is calculated as:

$$W = \frac{(\beta_1)^2}{Var(\hat{\beta}_1)}$$

Where:
$\hat{\beta}_1$ is the estimated coefficient for X,
Var(β_1) is the estimated variance of β_1

Distribution under Null Hypothesis
Under the null hypothesis (H0), the Wald statistic follows a chi-squared distribution with 1 degree of freedom.

Decision Rule
Compare the calculated Wald statistic to the critical value from the chi-squared distribution. If the calculated statistic is greater than the critical value, reject the null hypothesis, suggesting joint significance of the coefficient β_1

Interpretation
Rejecting the Null Hypothesis
This indicates that the slope coefficient β_1 is significantly different from zero, indicating a significant relationship between the independent variable (X) and the dependent variable (Y).

Failing to Reject the Null Hypothesis
The slope for the coefficient indicates joint significance, indicating that there is no evidence of a significant relationship between (X) and (Y).

R program for the Sup-Wald Test
This example illustrates how the Sup-Wald test can be used to estimate the joint significance of a single coefficient in a simple linear regression model.

Below is a simple R program that conducts a Sup-Wald Test for the joint significance of the coefficient in a simple linear regression model. This code generates 500 data points, fits a simple linear regression model, and performs a Sup-Wald Test as follows:

```
# Set seed for reproducibility
set.seed(123)

# Define number of data points n
 <- 500

# Generate example data
X <- rnorm(n)
Y <- 2 * X + rnorm(n)

# Fit a simple linear regression model lm_model
<- lm(Y ~ X)

# Print the summary of the linear model
print(summary(lm_model))

# Extract coefficients
beta0_hat <- coef(lm_model)[1] beta1_hat <-
coef(lm_model)[2]

# Calculate residuals residuals <- resid(lm_model)

# Calculate RSS (Residual Sum of Squares) RSS <-
sum(residuals^2)

# Calculate variance of beta1_hat X_bar <- mean(X)
var_beta1_hat <- RSS / ((n - 2) * sum((X - X_bar)^2))

# Calculate the Wald statistic
Wald_statistic <- (beta1_hat^2) / var_beta1_hat

# Degrees of freedom for the Wald statistic df_num <- 1 #
One coefficient being tested

df_denom <- n - 2 # Degrees of freedom for residuals
```

Critical value for a 5% significance level (chi-squared distribution) critical_value <- qchisq(0.95, df_num, lower.tail = FALSE)

Compare Wald statistic to critical value p_value <- 1 - pchisq(Wald_statistic, df_num)

# Print results
cat("Wald Statistic:", Wald_statistic, "\n") cat("Critical Value:", critical_value, "\n") cat("P-value:", p_value, "\n")

# Test the hypothesis at a 5% significance level if (Wald_statistic > critical_value) {
cat("Reject the null hypothesis. There is evidence of joint significance.\n")
} else {

cat("Fail to reject the null hypothesis. No evidence of joint significance.\n")
}

This script generates example data, fits a simple linear regression model, calculates the Residual Sum of Squares (RSS), and performs a Sup-Wald Test for the joint significance of the coefficient. The results include the Wald statistic, critical value, p-value, and a decision about rejecting or failing to reject the null hypothesis. Adjust the script according to your specific data and requirements.

III. RESULTS

Call:
lm(formula = Y ~ X)

Coefficients:
(Intercept) X
-0.0004678 1.9460271

Wald Statistic: 1751.059
Critical Value: 0.00393214
P-value: 0
Reject the null hypothesis. There is evidence of joint significance.

IV. CONCLUSION

We reject the null hypothesis and conclude that the slope coefficient is significantly different from zero, which indicates a significant relationship between the independent variable (X) and the dependent variable (Y).
In conclusion, the Sup-Wald test serves as a crucial tool in assessing the significance of coefficients within regression models, offering practical implications for empirical research and data analysis. By effectively bridging the gap between theoretical concepts and practical applications, this study contributes to the advancement of statistical techniques in regression analysis. Researchers can leverage the Sup-Wald test to enhance the rigor and credibility of their findings, ultimately fostering a deeper understanding of variable relationships and driving insights in various fields of study.

REFERENCES

1. Andrews, D. W. K. (1993): Tests for Parameter Instability and Structural Change With Unknown Change Point. Econometrica, 61, pp. 821-856.
2. Wenger, K. and Leschinski, C. and Sibbertsen, P. (2018): Change-in-mean tests in long-memory time series: a review of recent developments. AStA Advances in Statistical Analysis, 103:2, pp. 237-256.
3. Shao, X. (2011): A simple test of changes in mean in the possible presence of long- range dependence. Journal of Time Series Analysis, 32, pp. 598-606.
4. Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.
5. Bai, J. & Perron, P. (1998) Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.
6. Bai, J. & Perron, P. (2003) Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18, 1–22.
7. Barsky, R.B. (1987) The Fisher hypothesis and the forecastibility and persistence of inflation. Journal of Monetary Economics 19, 3–24.
8. Berk, K.N. (1974) Consistent autoregressive spectral estimates. Annals of Statistics 2, 489–502.
9. Burdekin, R.C.K. & Siklos, P.L. (1999) Exchange rate regimes and shifts in inflation persistence: Does nothing else matter. Journal of Money, Credit and Banking 31, 235–247.
10. Busetti, F. & Taylor, A.M.R. (2004) Tests of stationarity against a change in persistence. Journal of Econometrics 123, 33–66.
11. Harvey, D.I., Leybourne, S.J., & Taylor, A.M.R. (2006) Modified tests for a change in persistence. Journal of Econometrics 134, 441–469.
12. Kang, K.H., Kim, C.J., & Morley, J. (2009) Changes in U.S. inflation persistence. Studies in Nonlinear Dynamics & Econometrics vol. 13(4), article 1.
13. Kejriwal, M. & Perron, P. (2010) A sequential procedure to determine the number of breaks in trend with an integrated or stationary noise component. Journal of Time Series Analysis 31, 305–328.
14. Kim, D. & Perron, P. (2009) Unit root tests allowing for a break in the trend function under both the null and alternative hypotheses. Journal of Econometrics 148, 1–13.Google Scholar

15. Kim, J.Y. (2000) Detection of change in persistence of a linear time series. Journal of Econometrics 54, 159–178.Google Scholar
16. Kim, J.Y. (2003) Inference on segmented co integration. Econometric Theory 19, 620–639.
17. Mankiw, N.G., Miron, J.A., & Weil, D.N. (1987) The adjustment of expectations to change in regime: A study of the founding of the Federal Reserve. American Economic Review 77, 358–374.
18. Ng, S. & Perron, P. (2001) Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519–1554.
19. Perron, P. (1989) The great crash, the oil price shock, and the unit root hypothesis. Econometrica 57, 1361–1401.
20. Perron, P. & Qu, Z. (2006) Estimating restricted structural change models. Journal of Econometrics 134, 373–399.