Hermes-Model: Universal Model to Estimate The Rebuilding Costs of Houses

Michiel JELLEMA and Ashwan AUTAR, The Netherlands

Keywords: Valuation, Econometric Model, Real Estate, Hermes model, Data Quality

SUMMARY

The Dutch Association of Insurers (Verbond van Verzekeraars) has developed the so-called 'Herbouwwaardemeter Woningen'. This 'Herbouwwaardemeter' is a non-binding easy to use rebuild calculator for insurers, representatives and intermediaries to get an approximate estimate of the rebuilding costs of houses. Insurers make frequent use of this rebuild calculator or they have developed – sometimes based on this rebuild calculator– their own models.

In discussions with insurers, it became clear that the rebuild calculator and related models presently in use do not function optimally. They often lead to unreliable results, and with it, financial consequences for both the insurer and the customer. This is confirmed by research conducted on these models by Infofolio. Estimates of the rebuilding costs made by these models can differ from each other by 50 percent for exactly the same house.

Infofolio has developed the Hermes-model in cooperation with the Erasmus University Rotterdam and in conjunction with various insurers and subsequently has tested the model on theoretical criteria. The Hermes-model consists of three models, one main model and two submodels. The main model can be applied to regular houses up to a volume of 1.000 m3. The submodels are specialized models for houses that have a volume greater than 1.000 m³ and for houses registered as monument.

The analysis as described in this article shows that the Hermes-model is an objective and robust model that possesses high explanatory power to estimate the rebuilding costs of individual houses. The Hermes-model classifies the reliability of the individually calculated rebuilding costs into five categories.
Infofolio has used the Hermes-model to estimate the rebuilding costs for more than 7.3 million houses in the Netherlands and has analyzed the reliability of the model. Regarding the reliability of the estimated rebuilding costs for the 7.3 million houses, results show that 97,5 percent of the estimates deviate 15 percent or less from the actual rebuilding costs (status October 2014).

The rebuilding costs and its accompanying reliability, both estimated by the Hermes-model, together with additional data on these houses are used by insurers, representatives and intermediaries in different business processes. Nowadays the Hermes-model is applied to the offering process, the periodical check and renewal of policies, risk management, reinsurance, the execution of compliance and the adherence to the European Union Solvency II Directives (Jellema, 2013).

Hermes-Model: Universal Model to Estimate The Rebuilding Costs of Houses

Michiel JELLEMA and Ashwan AUTAR, The Netherlands

1. INTRODUCTION

An accurate estimate of the rebuilding costs of a house is an important part of an insurance policy. Research by Infofolio shows that many houses in the Netherlands are underinsured due to incorrect estimates of the rebuilding costs. Because of this, many insurers provide an additional guarantee against underinsurance (e.g. United Insurance). Both the insurer and the customer have a need for a highly reliable estimate of the rebuilding costs of the house to be insured.

Rebuilding costs are not strictly linked to the market value of a house. Rebuilding costs only take into consideration the cost of materials and labour that would be involved in rebuilding a house. The market value (the house and the land), on the other hand, takes into account other less tangible factors such as location and supply and demand.

To solve this problem Infofolio has developed the Hermes-model in cooperation with the Erasmus University Rotterdam and in conjunction with buildings insurers. The Hermes-model is an objective, unambiguous and robust econometric model which estimates the actual rebuilding costs of individual houses with a very high level of explanatory power. The Hermes-model also calculates the level of statistical reliability of the estimated rebuilding costs.

This article discusses the theoretical background and the methodology of the Hermes-model. It also describes the way the Hermes-model is tested on its theoretical robustness and the results of this test. The theoretical part of this paper is mainly quoted from the book Heij C. et al, 2004, Econometric Methods with Applications in Business and Economics. All tests and formulas used in this paper can be found in the appendix.
Finally, the article shows results of the Hermes-model when applied to the Netherlands as a whole.

2. DATA ACQUISITION AND TREATMENT

For the development of every econometric model – the Hermes-model being no exception – it is important to have access to relevant, reliable, objective and up-to-date data. During the development of the Hermes-model it was decided to use a wide variety of different housing data that are available on a nationwide scale. This data comes from several sources, namely national, provincial and local authorities and market parties.

Next - in order to maximize intersubjectivity – the actual rebuild values of houses were obtained from various insurers (www.infofolio.nl).

In order to prevent the use of certain regressors showing a strong relation with the rebuilding costs based on pure coincidence – and therewith not being representative for the whole population – only those regressors have been used in the development of the Hermes-model where an actual practical relation was foreseen in advance (Verbeek, 2012). The variables used in the development of the Hermes-model are housing data which are logically characterized in architectural, utilization, economic and geographical dimensions (Brouwer, 2007).

The sample on which the regressions are carried out are called the reference houses. All the data of these houses have been checked for accuracy and correctness.

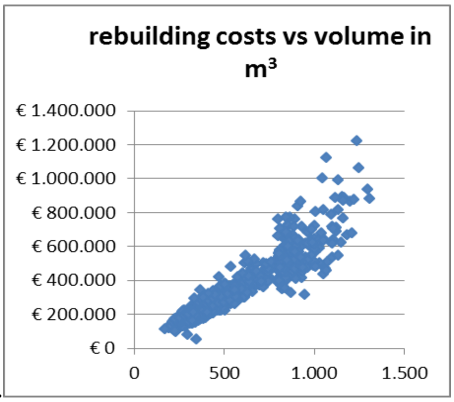
3. THE APPLIED METHODOLOGY

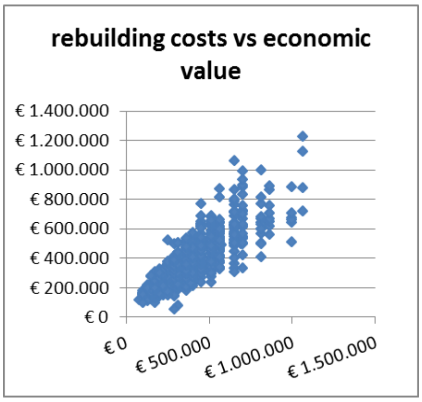
The Hermes-model consists of three models, one main model and two submodels. The main model can be applied to regular houses up to a volume of 1.000 m^3. The submodels are specialized models for houses that have a volume greater than 1.000 m^3 and for houses registered as monument. The reason for this division is that the margin of error in the estimate of rebuilding costs for big houses is larger and it is costlier for houses registered as monuments to rebuild.

For the development of these three different models, samples of regular, big and houses registered as monument were used respectively. The same methodology is used for their development. The methodology is described in the next paragraph, with the focus on the explanatory power and reliability of the Hermes-model.

3.1 Multiple linear regression analysis

The scatter charts below show a possible linear relation between the regressors (explaining or independent variables) and the rebuilding costs (dependent variable).





The volume and the economic value of the house are two of many variables used to estimate the rebuilding costs. Given the linear relation the choice was made to test the significance of the variables (Appendix (6)) and compute the explanatory power (Appendix (2)) of the regression with a multiple linear regression analysis, based on the Least Square Method (Appendix (1)) (Wackerley et al, 2008).

Besides the test of the individual regressors, the joint significance of the model was also tested with the F-test (Appendix (7)). The Variance Inflation Factor (VIF) has been used to measure the severity of correlations between independent variables (multicollinearity) (Appendix (3)) (Heij et al, 2004).

3.2 Testing seven assumptions

There are seven theoretical assumptions that justify the use of multiple linear regression analysis. Violation of any of these assumptions may have consequences for the reliability of the predictions made by the regression model. Ideally a model meets all seven assumptions, but in practice this rarely is the case. The application of the model indicates the importance of the seven assumptions. For example the test on serial correlation is very important when the model is based on time series data. Heij et al (2004) concludes that a model can also give reliable results when some assumptions are not satisfied. When assumptions are violated it is important to make adjustments in order to create the best performing model.

These are the seven assumptions:

1. Fixed independent variables: All elements of the data set have been obtained in a non-stochastic way.
2. Random error terms; the mean of the probability distribution of the error terms is zero.
3. Homoscedasticity: the variance of the error terms is constant in the whole data set. The Goldfeld-Quandt Test (Appendix (8)) is used to analyze whether the value of the error terms increase with higher estimates of the dependent variable.
4. No serial correlation; the error terms are not correlated.
5. Constant parameters: the parameters β (estimated coefficients) and σ (standard deviation) are constant over the whole dataset. The occurrence of varying parameters is tested with the CUSUM Test (Appendix (9)).

6. Linear model: the dependent variable y is generated by the linear model $y = X\beta + \varepsilon$, with β as estimated coefficients and ε the error terms. The misspecification of the linear model is tested with the Ramsey RESET Test (Heij et al, 2004).
7. Normality: the probability distribution of the error terms is a normal distribution. The normality of the error terms is tested with the Jarque-Bera Test (Appendix (10)).

3.3 Providing a margin of error

For every individually calculated rebuilding costs the Hermes-model gives an indication of its reliability by providing a margin of error. The margins of error are classified into one of five categories and every category is represented by a number of stars (see table below).

Number of stars	Margin of error
5	< 10%
4	10% - 15%
3	15% - 20%
2	20% - 25%
1	> 25%

The margin of error depends on the quality of the data used to estimate the rebuilding costs. Input mistakes, measurement errors or timeliness errors lower the quality of the data thereby lowering the number of stars. An estimate of the rebuilding costs based on high quality data will receive a high number of stars, signifying a more reliable estimation.

The quality of the data is determined by comparing the data of the house for which the rebuilding costs are estimated with very high quality data of the reference houses (appendix (4) and (5)) (Poole, 2011 and Heij et al, 2004).

The formula (appendix (5)) calculates for each house a reliability factor $\sqrt{d-1}$ and a percentage of the deviation based on the combined quality of the data. The higher this factor and the higher this percentage of the deviation, the lesser the actual values of the data correspond with those of the house selected as reference. This results in a less reliable estimate of the rebuilding costs. In these cases the Hermes-model gives fewer stars to the estimated rebuilding costs.

4. RESULTS AND CONCLUDING REMARKS

4.1 Regression analysis

The regression analysis shows that not all variables are significant, these variables are omitted from the model. The results also shows that some variables are more or less significant than others. For example house age is less significant than volume. Outliers can have a negative effect on the explanatory power of the model. Removing these outliers results in better estimates and higher explanatory power (Paap, 2001).

The result of the root of the Variance Inflation Factor indicates that multicollinearity exists to a small extent but does not hamper the development of the Hermes-model.

4.2 Seven assumptions

The Hermes-model meets the first, second and fourth assumption in advance.

The first criterion is met, because all elements of the data set used in developing the Hermes-model are fixed independent variables. The second criterion is met, because the application of a multiple linear regression analysis implies by definition the mean of the probability distribution of the error terms is zero. The fourth criterion is also met, because the used variables are not time-dependent.

The remaining assumptions were tested and results show that the main model and the submodel for monuments meet all assumption except assumption 7. The submodel for big houses does not meet assumption 3 and 7.

4.3 Prediction

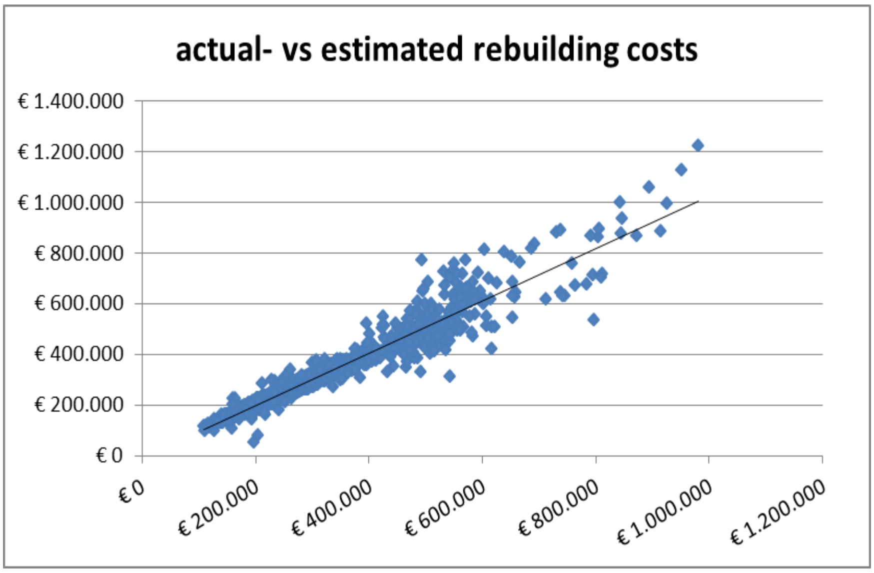
As a final test, unrelated to the seven assumptions discussed above, the Chow Forecast test (Heij et al, 2004) was conducted. This test shows that the rebuilding costs prediction errors are acceptable.

4.4 Margin of error

The use of the equation formula of the margin of error has been tested. As a test some data was changed on purpose and these changes directly led to a different outcome of the reliability. For instance, a significant change in floor area of a house directly led to a change in reliability by a number of stars.

4.5 Explanatory power

Based on these tests the results show that the main model is a robust model which estimates rebuilding costs of regular houses in the Netherlands with 95 percent explanatory power (R^2 = 0,95). The explanatory power of the submodel for big houses is 87 percent and the submodel for monuments has 89 percent explanatory power. A scatter diagram of the actual- versus the estimated rebuilding costs is shown below.



4.6 Concluding remarks

One of the problems faced by insurers is to determine which rebuilding cost is calculated accurately and which not, due to input data which could be unreliable. The developed model solves this problem by providing a margin of error for each estimated rebuilding cost based on the input variables. The Hermes-model has an explanatory power of 95% and can be applied to all five types of houses of all sizes. This means that the input variables “explain” the rebuilding costs for 95%. This accuracy is a strong improvement compared to the current models.

This research is aimed at developing an objective and robust model that possesses high explanatory power to estimate the rebuilding costs of individual houses. Based on the described tests and the reliability formula the conclusion can be drawn that the developed Hermes-model is a robust, objective and unambiguous model which estimates rebuilding costs accurately and provides a margin of error.

5. THE HERMES-MODEL IN PRACTICE

Infofolio has applied the Hermes-model to more than 7.3 million houses (Centraal Bureau voor Statistiek) in the Netherlands and has analyzed the reliability of the estimated rebuilding costs (status October 2014). In this chapter the use and results of the Hermes-model will be described from different points of view.

5.1 Use of Hermes-model

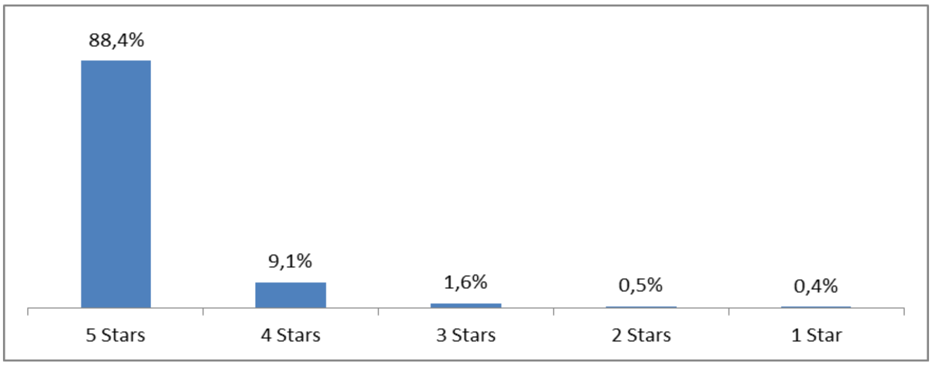
Rebuilding costs and their reliability estimates, both calculated by the Hermes-model, together with other data on the 7.3 million houses are nowadays frequently used by more than 50 insurers, representatives and intermediaries in different business processes.
The Hermes-model is used in the offering process, the periodical check and renewal of policies, risk management, reinsurance, the execution of compliance and the adherence to the European Union Solvency II Directives.

5.3 Reliability of the rebuilding costs

The Hermes-model works best for rebuilding costs between € 100.000 to € 850.000. Almost all houses fall into this category.

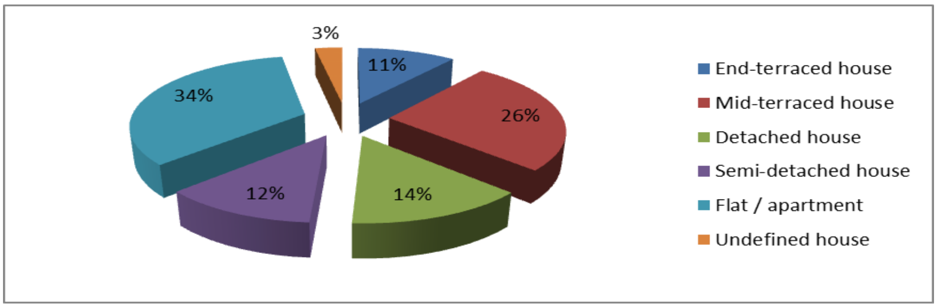
As described in this article, every estimated rebuilding costs gets a reliability classification expressed by a number of stars. Four and five star estimates are always used by insurers, representatives and intermediaries.

The distribution of reliability classifications (stars) of the 7.3 million houses is as follows;



5.2 Houses by category

These 7.3 million houses can be categorized as follows;



5.4 Continuous mutations in rebuilding costs

Every two months the estimates of the rebuilding costs are recalculated based on new and actualized housing data. And every six months the rebuilding costs are updated with a regional price-index.

Over the year 2013 this continuous updating-process resulted in the following mutations;
- 80% of all the rebuilding costs changed with a maximum of +/- 5%
- 90% of all the rebuilding costs changed with a maximum of +/- 10%
- 10% of all the rebuilding costs changed with more than +/- 10%.

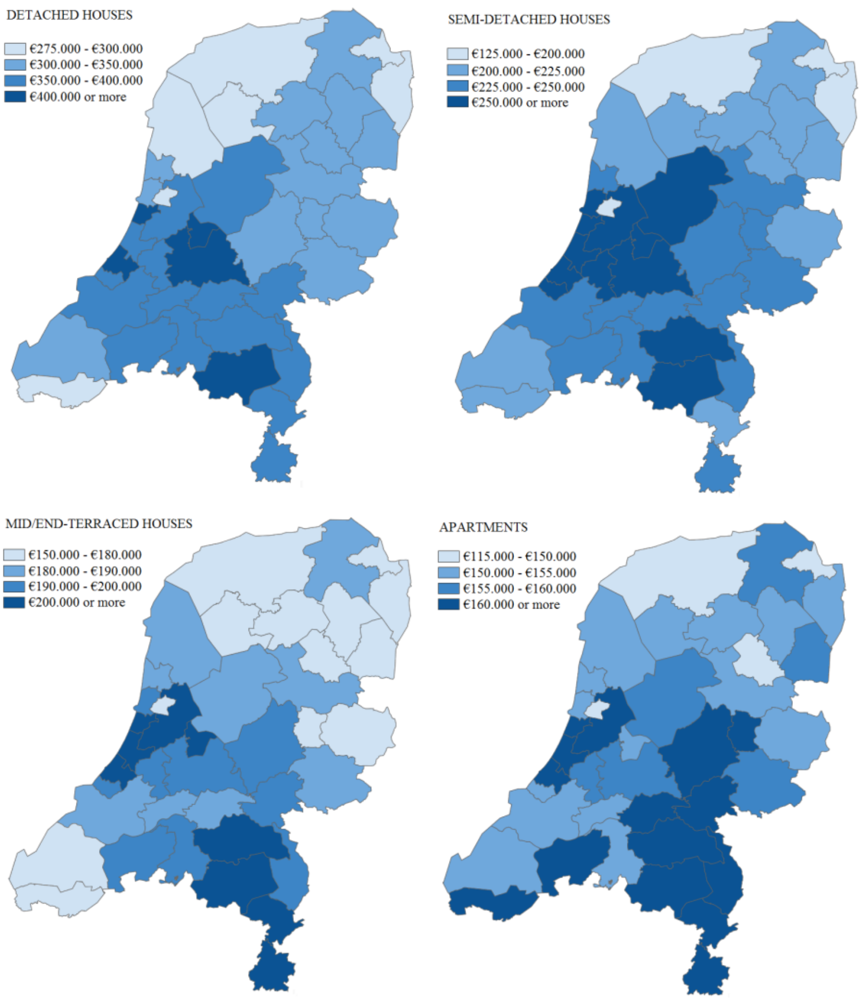
All this mutations lead to the following notions:
- The average rebuilding costs decreased over the year 2013 by -0.4%
- 37% of the rebuilding costs of all houses increased by an average of 5.7%
- 62% of the rebuilding costs of all houses decreased by an average of 4.2%
- 1% of the rebuilding costs of all houses did not change.

5.5 Regional differences in rebuilding costs

The average rebuilding costs of all houses in the Netherlands (status October 2014) are for:

- Detached houses	€ 349.300
- Semi-detached houses	€ 237.900
- Mid / End-terraced houses	€ 192.600
- Flat / apartment	€ 157.500

Due to regional differences in the costs of materials, labour and other local prices the rebuilding costs of houses differs from region to region. The next figure shows the regional differences in rebuilding costs for detached houses, semi-detached houses, mid / end-terraced houses and apartments.



5.6 Development of the Hermes-model for non-residential buildings

The methodology of the Hermes-model is used for further development of other models. Since autumn 2014 the Mercurius-model is available. The Mercurius-model estimates the rebuilding costs of 400.000 non-residential buildings like offices, restaurants and stores.

REFERENCES

- Brouwer R., 2007, De Baten van Wonen aan Water: Een Hedonische Prijsstudie naar de relatie tussen Huizenprijzen, Watertypen en Waterkwaliteit, Vrije Universiteit Amsterdam, Amsterdam.
- Centraal Bureau voor Statistiek, www.cbs.nl
- Heij C. et al, 2004, Econometric Methods with Applications in Business and Economics, Oxford University Press, Oxford.
- Infofolio. www.infofolio.nl
- Jellema M., 2013, Hermes-model: Robuust model voor bepaling van de herbouwwaarde van woningen, De Actuaris, AG&AI, Utrecht.
- Paap R., 2001, Quantitative Models in Marketing Research,Cambridge University Press, Cambridge.
- Poole D., Lineair Algebra: a modern introduction, 2011, Richard Stratton, Boston.
- Verbeek M., A guide to modern econometrics, 2012, Wiley, New Jersey.
- Verbond van Verzekeraars, www.verzekeraars.nl
- Wackerley D. et al., Mathematical statistics and probability, 2008, Thomson-Brooks/Cole, California.
- United Insurance, www.unitedinsurance.nl, https://www.unitedinsurance.nl/garantie-tegen-onderverzekering-wat-is-dat.php.

BIOGRAPHICAL NOTES

Michiel Jellema has over 30 years of experience in the geo-information sector. He did this in different positions and roles but always bridging organizational, Geo-ICT and other aspects of making geo-information widely accessible. Since 2007 he is owner and director of Infofolio B.V.. In 2013 he earned his Phd-degree at the Delft University of Technology.

AshwanAutar is a student of Econometrics at the Erasmus University Rotterdam. He has 5 years of experience in the field of data analysis and econometric modelling.

CONTACTS

Dr. ir. Michiel Jellema
Infofolio B.V.
P.O. Box 72
3700 AB Zeist
THE NETHERLANDS
Email: m.jellema@infofolio.nl
Web site: www.infofolio.nl

Ashwan Autar
Erasmus Universiteit Rotterdam
Burgemeester Oudlaan 50
3062 PA Rotterdam
THE NETHERLANDS
Email: 266880aa@eur.nl
Web site: www.eur.nl

APPENDIX

The abstract formula of the model is shown in equation (1) where y represents the rebuilding costs, X the regressors, β the estimated coefficients and ε the error terms.

$$y = X\beta + \varepsilon \tag{1}$$

The explanatory power of the model is measured by the coefficient of determination R^2 which is defined as:

$$R^2 = \frac{SSE}{SST} = \frac{b'X'NXb}{y'Ny} \text{ with } N = I - \frac{1}{n}ii' \text{ and } b = (X'X)^{-1}X'y \tag{2}$$

Variance Inflation Factor (VIF):

$$VIF = \frac{1}{1-R_j^2}, \tag{3}$$

$with\ R_j^2\ the\ coefficient\ of\ determination\ of\ the\ regression\ of\ regressor\ j\ on\ the\ rest\ of\ the\ regressors$

The method to calculate the margin of error per house is shown in the equations (4) and (5). In equation (8) I is the identity matrix, $(nxk)\ matrix\ X_1$ the regressors (with constant term) of the reference houses and $(gxk)\ matrix\ X_2$ the regressors (with constant term) of the prediction sample.

$$M = I + X_2(X_1'X_1)^{-1}X_2' = \begin{bmatrix} a_{11} & \cdots & a_{1g} \\ \vdots & \ddots & \vdots \\ a_{g1} & \cdots & a_{gg} \end{bmatrix} \tag{4}$$

$$d = diag(M) \rightarrow margin\ of\ error = \sqrt{d-1} * 100\% \tag{5}$$

<u>Statistical tests</u>

A rejection region of five percent is used in all tests.

T-test: $t_j = \frac{b_j}{s_j} \sim t(n-k),$ (6)

$with\ s_j\ the\ standarddeviation\ of\ regressor\ j, n = sample\ size, k = number\ of\ regressors$

F-test: $F = \frac{n-k}{k-1}\frac{R^2}{1-R^2} \sim F(g, n-k),\ with\ g = k-1$ (7)

Goldfeld-Quandt Test with test statistic:

$$F = \frac{s_2^2}{s_1^2} \sim F(n_2 - k, n_1 - k) \tag{8}$$

CUSUM Test with test statistic: $W_r = \sum_{t=k+1}^{r} \frac{w_t}{s} \sim N(0, r-k)$ (9)

Jarque-Bera Test with test statistic: $JB = n(\frac{1}{6}S^2 + \frac{1}{24}(K-3)^2) \sim \chi^2(2)$. The normal distribution has a low JB-value, a skewness (S) of 0 and a kurtosis (K) of 3. (10)