MACQUARIE University

Business Forecasting

Optional Advanced Notes on Regression

Checking Regression Assumptions

MACQUARIE University

Before using any regression models for interpretation or forecasting you need to **ensure the assumptions of the regression model are valid**

Requires checks of model adequacy involving **model significance tests** and the **residuals of the model**

This includes visual inspection of **residual plots** against **time** and against **predicted Y** and **all X's**

There will also be several **statistical tests** on the **residuals** to further test the validity of error assumptions

Common Problems with Regression

MACQUARIE University

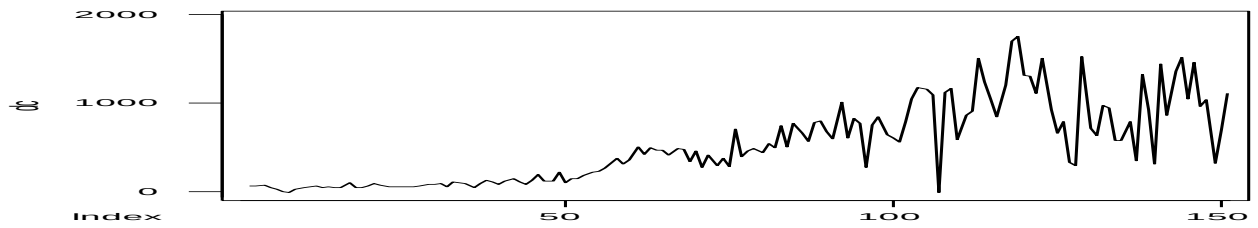
Model Mis-specification: Exclusion of relevant variables, inclusion of irrelevant variables and incorrect functional form. Mis-specified models likely to lead to **erroneous conclusions** and **forecasts**. Detection: Logic, economic theory, model significance tests and residual plots. Remedy: Use correct model

Heteroskedasticity: Constant variance assumption of the error term is violated. Commonly, **residuals increase in magnitude over time**

Violation of homoscedasticity (constant error variance) may lead to **inefficient estimates and forecasts** (still **unbiased**) and **invalidity of the standard inferential tests**. Detection: White Test and residual plots. Remedy: use logarithms of Y, X

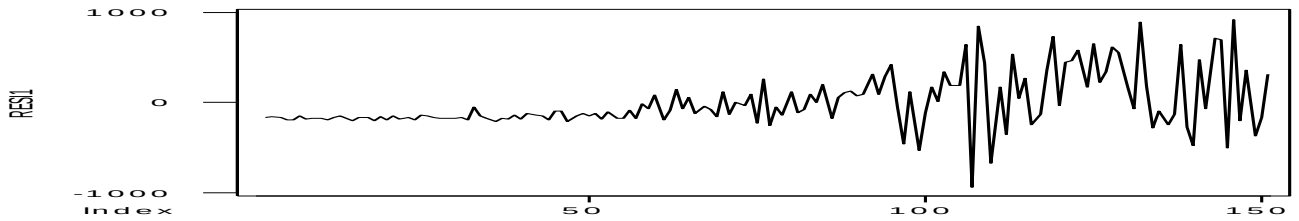
Heteroscedasticity

Detection: Mostly evidence from **time series plots** will establish if there is heteroscedasticity.



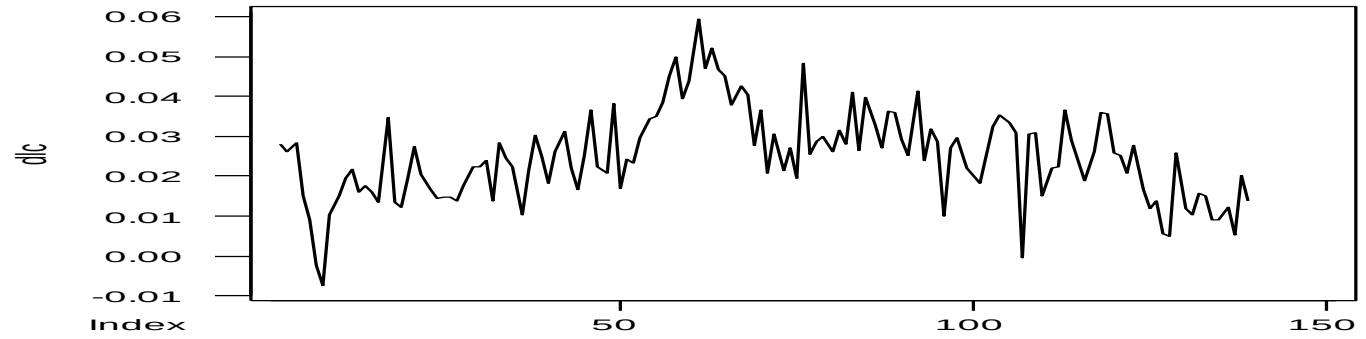
From the example above, **volatility of the time series (random component variance)** is increasing over time.

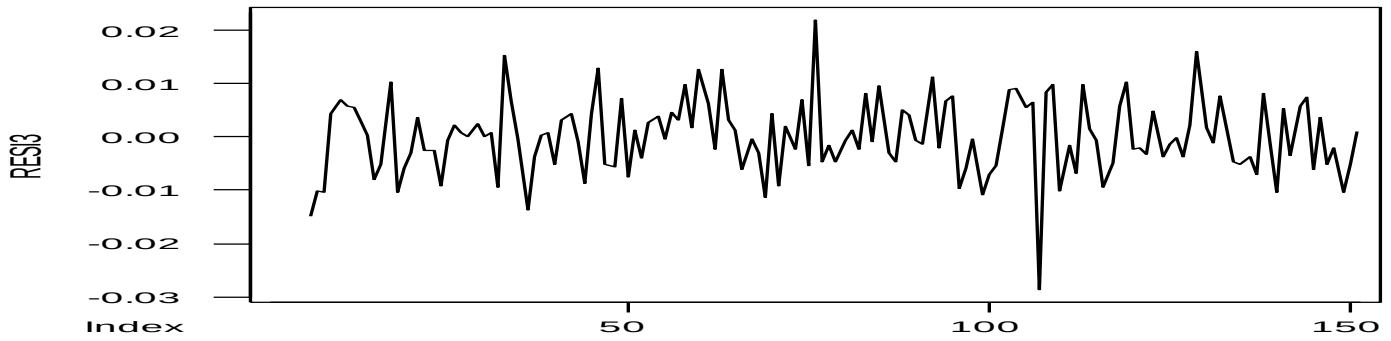
Alternatively, **plots of residuals** from estimated regression model may also show heteroscedastic nature of the time series



Remedy (example)

Below are **logarithmic transformations** (natural logs) of the previous graphs.





Autocorrelation

MACQUARIE University

A key assumption for **OLS estimation efficiency** is **independence** of the error term with respect to error terms at other observations

May be violated with time series data (**autocorrelation)**

Error in one period may be related to error terms in other periods

$$\varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\ldots\ldots\ldots \varepsilon_{t-p})$$

Severe autocorrelation may lead to **inefficient and/or biased estimation/forecasts**

Detection: Residual ACF and PACF, DW test (1st order), LM test (any order)

Remedy: Correct specifications, Autocorrelation corrections

Time Series Regressions

Time Series regressions present specific estimation problems due to the likely **connection** between observations over time

For example, **two unconnected (in theory) time series that are both trending** or have similar structure are likely to produce an **apparently significant** relationship in a regression even though in theory (and in practice) they are not connected

In general, problems in **estimating and interpreting regressions** **may** arise when the levels of the series used (Y and X's) vary with time or the time series are non-stationary

This is called spurious regression and can lead to **erroneous conclusions and forecasts**

Spurious Regression

MACQUARIE University

Regression of non-stationary variables may lead to **erroneous estimation results**

Experiments show two time series that are **non-stationary** but are **independently generated** (ie no relationship exists between the time series) if regressed against one another **may** lead to results which suggest the variables are related **(large t and F statistics, R^2 close to 1)**

The relevant regression tests and inferential procedures will suggest incorrectly **the two variables have a relationship and the estimated regression model is valid**

Forecasts and predictions are likely to be poor

Spurious Regression (cont)

MACQUARIE University

The consequences of spurious regression results cannot be overstated

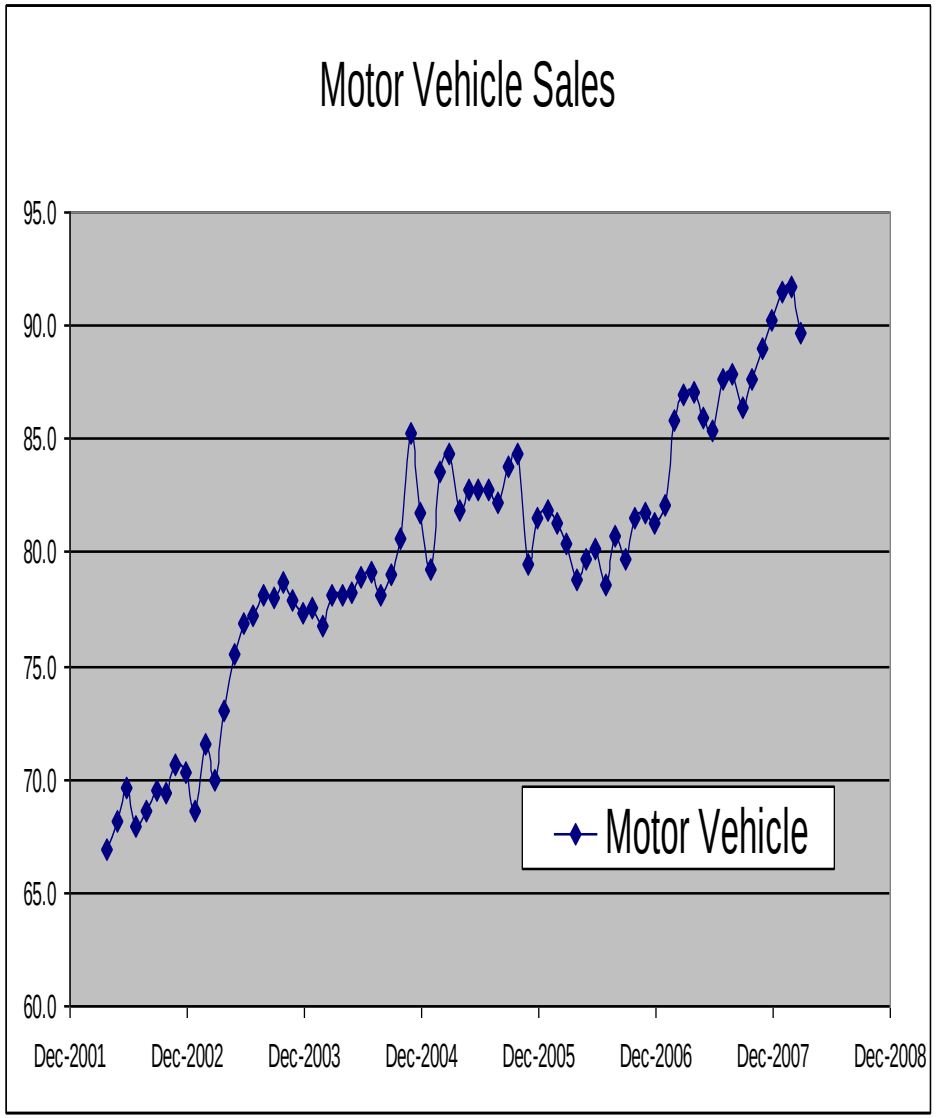
Results will suggest the model will be a useful predictive tool (R^2 high, F high). Yet, since **no relationship** exists between the variables **predictive performance will be poor**

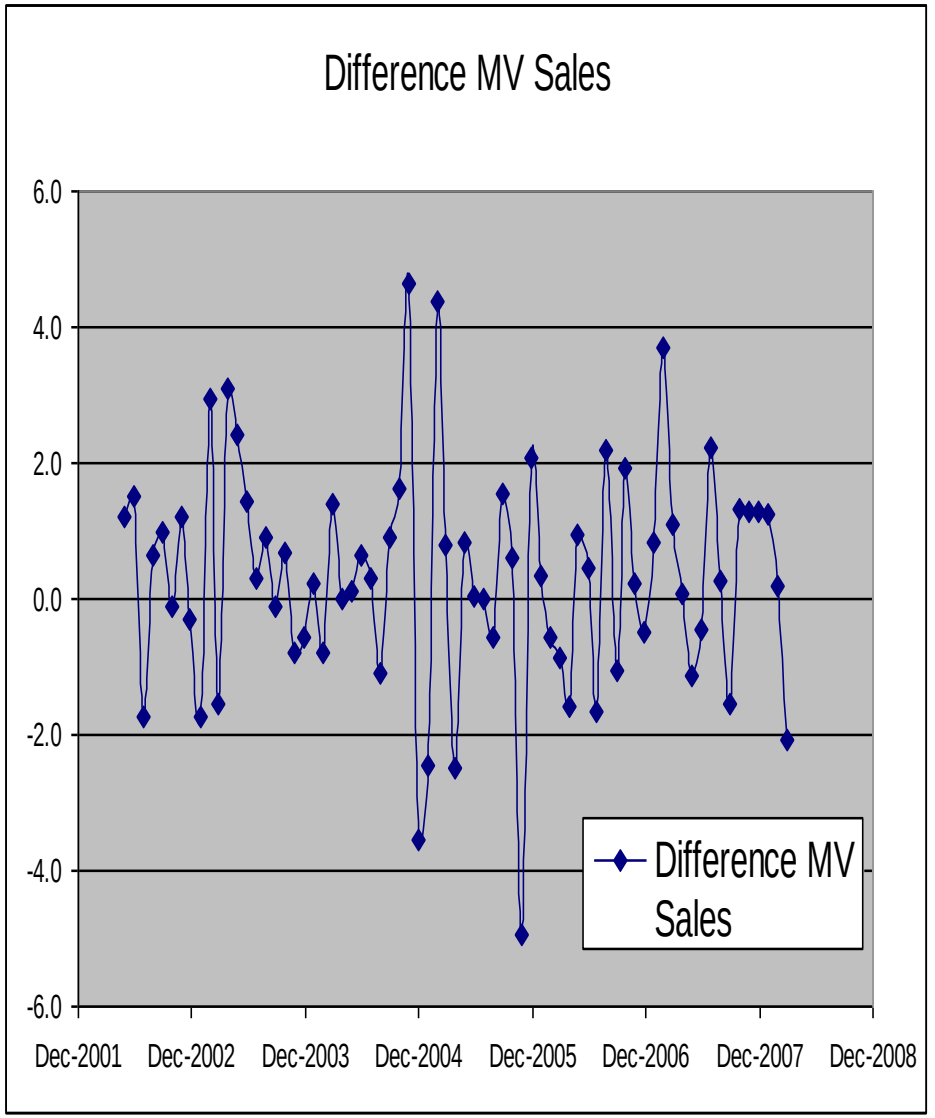
Therefore great caution must be exercised in regression of time series to ensure **estimated results are valid**

Remedy: Economic Theory, Logic, Co-integration, Use Stationary series (possibly difference original time series)

Detection: Line Plots, ADF test (Non-stationarity)

Comparing Original and Differenced





Dynamic Structure

MACQUARIE University

It is likely impacts of independent variables on the dependent variable will not be entirely **instantaneous**

The modeller needs to consider a **dynamic structure**

The typical dynamic models used are the **distributed lag models** or models using **lagged dependent variables** or models with both

Selection of the appropriate dynamic model is a complex task with many pitfalls for the unwary. There may possibly be **estimation problems or theoretical problems** with any type of dynamic structure chosen. It is best to consult an expert.