Macroeconomic Determinants of Real Estate Returns

An econometric analysis of the macroeconomic influence on the United States Commercial Real Estate returns

THOMAS KOFOED-PIHL

Applied Economics and Finance
MSc in Economics and Business Administration

Supervisor: Lisbeth La Cour
Department of Economics


…real estate combines sociology, geography, demography, architecture, and political forces, with the dynamics of fundamental economic trends, complex financing problems, the perils of illiquidity, and subtle valuation considerations.

Peter L. Bernstein
Table of Content

1 INTRODUCTION ............................................................................................................. 9
   1.1 PROBLEM STATEMENT ................................................................................................. 11
   1.2 METHODOLOGY ........................................................................................................... 12

2 MODEL AND DATA CHOICE ..................................................................................... 13
   2.1 CHOICE OF REAL ESTATE RETURN TIME SERIES ............................................................ 14
   2.2 CHOICE OF MACROECONOMIC TIME SERIES ................................................................. 14
   2.3 DEFINITION OF THE MODEL .......................................................................................... 15
   2.4 LITERATURE REVIEW ................................................................................................... 16

3 PROPERTY VALUATION AND UNSMOOTHING .................................................. 18
   3.1 NCREIF NATIONAL PROPERTY INDEX (NPI) .............................................................. 20
   3.2 MIT TRANSACTION BASED INDEX (TBI) ..................................................................... 21
   3.3 COMPARISON OF NPI AND TBI .................................................................................... 22
   3.4 NOISE AND APPRAISAL ERROR IN REAL ESTATE VALUATION ................................ 23
      3.4.1 Effects of the errors ............................................................................................. 25
   3.5 THE METHODOLOGY OF MOVING AVERAGES .......................................................... 27
   3.6 METHODS OF UNSMOOTHING APPRAISAL-BASED RETURNS ..................................... 28
      3.6.1 Zero-autocorrelation technique ........................................................................... 28
      3.6.2 Mechanical de-lagging (mathematical approach) ............................................... 29
      3.6.3 Regression modelling (econometric approach) ................................................... 33
   3.7 INTRODUCTION TO ECONOMETRIC ANALYSIS ...................................................... 34
4 STATIONARITY OF THE TIME SERIES ................................................................. 36

4.1 STATIONARITY TESTS FOR TBI ................................................................. 37
   4.1.1 Durbin-Watson d-statistic ........................................................................ 39
   4.1.2 The Breusch-Godfrey Series Correlation test (LM test) ......................... 40
   4.1.3 Correlograms of the ACF and PACF ....................................................... 41
   4.1.4 Dickey-Fuller Unit Root Test (DF-test) ..................................................... 43

4.2 STATIONARITY AMONG THE EXPLANATORY MACRO VARIABLES .......... 44

5 CO-INTEGRATION AMONG UNIT ROOT TIME SERIES .................................... 46

5.1 EXPECTATIONS AND PERFORMANCE OF CO-INTEGRATION TESTS ............ 47
   5.1.1 Co-integrating Regression Durbin-Watson (CRDW) Test ....................... 48
   5.1.2 Augmented Engle-Granger (AEG) Test .................................................... 49
   5.1.3 Pair wise co-integration tests of variables ................................................. 50

5.2 THE ERROR CORRECTION MODEL 1 ............................................................ 52

6 CLASSICAL LINEAR REGRESSION MODEL ASSUMPTIONS .......................... 54

7 MODEL SPECIFICATION ................................................................................... 56

   7.1 JARQUE-BERA NORMALITY TEST ............................................................... 58
   7.2 RAMSEY’S REGRESSION SPECIFICATION ERROR TEST ...................... 58
   7.3 OUTLIERS AMONG RESIDUALS ................................................................. 60
      7.3.1 Reasoning behind removal of outliers .................................................... 61
      7.3.2 Removal of outliers and model selection .............................................. 63
   7.4 THE ADJUSTED ERROR CORRECTION MODEL 2 ..................................... 65

8 TIME SERIES VOLATILITY ............................................................................... 68

   8.1 ARCH TEST ON MODEL 2 ............................................................................. 70
      8.1.1 Graphical indication of ARCH effects .................................................... 71
      8.1.2 Engle’s ARCH test ................................................................................. 71
      8.1.3 LM test for ARCH disturbances ............................................................ 73

9 AUTOCORRELATION ......................................................................................... 74

10 MODEL ESTIMATION ...................................................................................... 76
List of Figures

3.1 Real estate reservation price trade off (Source: Geltner, Miller et al. 2007)
3.2 Comparison of Transaction-Based Index (TBI) and NCREIF Property Index (NPI) (Source: Jim Clayton, PREA, 2008)
3.3 Real estate valuation noise-lag trade-off frontier (Source: Geltner, Miller et al. 2006)
3.4 Comparison of statistical errors of valuation (Source: Geltner, Miller et al. 2007)
3.5 Zero-Autocorrelation Technique (Based on Geltner, Miller et al 2007)
3.6 TBI Total Return Index (Source: MIT Centre for Real Estate)
3.7 Mechanical de-lagging (Own production)
4.1 Changes in log values of TBI (Own production)
7.1 Histogram and Probability plot of Error Correction Model 1 (SAS output)
7.2 Residual plot of Error Correction Model 1 (SAS output)
7.3 TBI time series – Error Correction Model 1 (SAS output)
7.4 TBI time series – Adjusted Error Correction Model. Model 2 (SAS output)
7.5 Histogram and Probability plot of Model 2 (SAS output)
7.6 Residual plot of model 2 (SAS output)
7.7 Residual plot for autocorrelation of model 2 (SAS output)
7.8 Scatter plot of log GDP versus Log INT (SAS output)
**List of Tables**

3.1 Classic Real Estate Indexes in the United States (Source: Giorgiev et al. 2003)
4.1 TBI Dubin-Watson test (SAS output)
4.2 TBI Breush-Godfrey Serial Correlation test (LM test) (SAS output)
4.3 TBI ACF and PACF Correlograms (SAS output)
4.4 TBI ACF and PACF Correlograms differenced once (SAS output)
4.5 TBI Dickey-Fuller Unit Root test (SAS output)
5.1 Co-integration test among the explanatory variables (SAS output)
7.1 Ramsey’s specification test of Model 1 (SAS output)
7.2 Error Correction Model selection via AIC and SIC (based on SAS output)
8.1 ACF and PACF of squared residuals from Model 2 (SAS output)
8.2 Model 2 ARCH test for higher-order p (SAS output)
9.1 Model 2 Breush-Godfrey Serial Correlation test (LM test) (SAS output)
10.1 Correlations of the log values of the variables (Own production)
Abstract

Real estate investments have begun to make an impasse in professional investors’ investment portfolios. Real estate is a rather new asset class, and for that reason the knowledge base of the asset is limited and it is less efficient and transparent compared to the stock and bond markets. In that view professional investors lack focus on the market fundamentals given by the development of the macro economy.

The aim of this paper is to examine whether the United States commercial real estate market is significantly influenced by changes in the macro economy. In doing so an error correction regression model is developed to econometrically analyse the macroeconomic determinants of the quarterly real estate total returns from 1984-2008. To address the smoothness problem of appraisal-based returns, the Massachusetts Institute of Technology’s unsmoothed transaction-based return index is applied to the regression model.

I found that unemployment and the long term interest rate are negatively influencing the total return of the US real estate market, while the gross domestic product over time heavily influences the total return positively. The fourth analysed variable inflation was found not to have any noticeable influence on the return.

The conclusions of the analysis reveal the importance of timing the real estate investments according to the development of the macro economy, and consequently that professional investors should focus on the development of the general economy rather than only on real estate specific characteristics.
Chapter 1
Introduction

Commercial real estate is a rather new asset class becoming a playground for numerous investors, corporations and managers that are looking for long term investments. Real estate is a challenging asset class due to its unique and heterogenic character, and it is typically traded between individual buyers and sellers making real estate an illiquid and non-transparent market. However, the asset also has wide-ranging qualities\(^1\), e.g. it is a good source of diversification, and a generator of attractive risk-adjusted returns through its low risk and high Sharpe Ratio. It offers opportunities of hedging against unexpected inflation and finally, real estate is regarded a strong cash flow generator through the income component of the return.

Real estate return is based on the two elements income and capital appreciation, the first refers to the housing rent, and the latter refers to the appreciation of the property value over time. Further, it can be divided into different sub-segments regarding type of investment and risk profile; Real estate professionals usually refer to these property types: (i) residential, (ii) retail, (iii) office and (iv) industrial plus a smaller sector of hotels and convention\(^2\).

The commercial property market accounts for roughly 8% of the investment universe\(^3\), however due to illiquidity and non-transparency of the market, research shows an average

\(^1\) Hudson-Wilson, Fabozzi, and Gordon, 2003, page 13 and 17-20
\(^2\) Geltner, Miller, Clayton and Eichholtz, 2007, page 112
\(^3\) Hudson-Wilson, Fabozzi, Gordon, 2004
The increased interest on commercial real estate in recent years is also mirrored in the academic literature, for instance *the Journal of Portfolio Management* published only six articles on the topic between the years of 1980 to 2003. Since then a special issue of the journal has been dedicated entirely to real estate each year.

The purpose of this paper is to investigate to what extent the macro economy influence the rate of return of the United States commercial real estate market. The topic caught my interest while working with international real estate investment during my masters’ studies. A real estate investor typically focuses a great deal on the location and quality of the asset, in order to build a foundation upon which they can achieve superior returns – but to what degree is the return simply a product of changes in the macroeconomic space?

The investment cycle of real estate is a constitution of the capital raising from investors, creation of value through acquisitions and active asset management and finally, the exit point, where the asset is being liquidated and a capital flow is generated back to the investors. The focus of this paper will be the value creation part. More precisely via a log-linear regression model I will analyse to which extent real estate market fundamentals, i.e. the macro economy, influence the value of US real estate returns.

---

6 In terms of development projects, refurbishments etc.
1.1 Problem statement

Inspired by my experience within real estate investment, and based on a global office market return analysis by De Wit and Van Dijk, the research question of this thesis is as follows:

Are the US real estate returns significantly influenced by the shifts in market fundamentals given by the changes in the macro economy over time?

Given the log-linear model fulfils the assumptions for the classical linear regression model; the research question will be examined through following hypotheses tests:

- The proxies for economic growth GDP and unemployment are, respectively, positively and negatively related to the real estate total return
- The real estate total returns are positively influenced by the level of inflation
- The proxy for money availability i.e. long-term interest rate is negatively related to the total returns in real estate markets
1.2 Methodology

The thesis is divided into a three-stage analysis of the macroeconomic determinants of real estate returns. An inductive research strategy is applied to the thesis implying the empirical dataset will be the foundation of the econometric analysis.

In stage one the real estate market characteristics were briefly introduced to the reader introduction-wise, in order to assure a basic understanding of the asset class. The purpose of this section is to highlight main qualities of the asset and presenting an overview of the real estate investment market.

In stage two the complications of property valuation issues are discussed and two approaches for valuation are described, namely the appraisal-based approach and the transaction-based approach plus the econometric implication associated hereto. Equivalently, two US real estate return indexes are presented using the different valuation approaches. Lastly, three methods of adjustment for smoothing processes in real estate returns are investigated before applying the most favourable one to the regression analysis.

Stage three consists of the actual empirical analysis through econometric modelling and regression of the real estate return index against the macroeconomic variables of unemployment, inflation, interest rate and GDP. The analysis is performed subject to the assumptions given by the classical linear regression model. Through hypothesis testing the development and estimation of the regression model will enable me to conclude on the thesis problem statement.

Finally, the findings of the analysis will be combined in an overall conclusion of the real estate returns determinants.
Chapter 2
Model and Data Choice

The thesis’ regression model of the macroeconomic variables influence on real estate return will take its offset in a model developed by De Wit and Van Dijk\(^7\) from the Dutch Real Estate Investment Company ING and ING Clarion, respectively. ING is among the biggest real estate players globally and ING Clarion is the US investment management and development arm of the ING real estate investment group\(^8\).

De Wit and Van Dijk investigated in 2007 the quarterly direct *office* real estate market returns from 1986 to 1999 across Asia, Europe and the United States, and the aim of the research paper was to identify the most important determinants of the changes in office returns and to understand the key implications for real estate investors\(^9\). The outcome of their research was a dynamic regression model including office-sector specific variables as vacancy and office stock and macroeconomic variables consisting of GDP, unemployment, inflation and interest rate. Prior to the model development the problem of smoothness in appraisal-based real estate returns was accounted for in the dynamic multivariate regression model.

The main findings of the global office market return analysis follow their prior expectations to the model. *The capital value* of office prices was significantly influenced by GDP, and negatively by the unemployment rate - whereas inflation had no influence. A lagged value of the return variable had only a modest effect. *The income component* also appears to be

\(^7\) De Wit and Van Dijk, 2007, *The Global Determinants of Direct Office Real Estate Returns*
\(^8\) [www.ing.com](http://www.ing.com) and [www.ingclarion.com](http://www.ingclarion.com)
\(^9\) De Wit and Van Dijk, 2007, page 29
negatively influenced by unemployment, while GDP and inflation both are insignificant regressed on the net rent of office returns.

On an aggregate level the outcome of the total return regression corresponds to those of capital appreciation and income. The unemployment rate is significantly negatively related to total return, while the parameter estimate of GDP is significantly positive. The change in inflation is however still insignificant on a total return base reflecting the inflation-hedge capabilities of real estate.

2.1 Choice of real estate return time series

The analysis made by De Wit and Van Dijk had a global perspective on the office market. This paper will narrow the perspective from the global scene and solely concentrate on the direct real estate market in the United States. That said however the analysis will not only take the office sector under consideration, but focus on an aggregate total return level across all real estate sectors. As mentioned before smoothing of appraisals-based returns can cause problems in analysing real estate data. For that reason the dependent variable of the regression model will consist of a transaction-based return index obtained via a hedonic pricing model from the MIT Centre of Real Estate\textsuperscript{10}. The complications of property valuation and the mechanisms to adjust for smoothing processes in real estate returns will be discussed prior to the empirical analysis.

2.2 Choice of macroeconomic time series

The regression analysis will examine four macroeconomic variables identical to those of the De Wit and Van Dijk paper, which are expected to influence the return base of the US real estate market. All data are withdrawn quarterly from DataStream in the period from Q1

\textsuperscript{10} MIT Centre of Real Estate Transaction-Based Index: http://web.mit.edu/cre/research/credl/tbi.html
1984 to Q1 2008. Following is a short description of the macroeconomic variables chosen:

1. GDP - a proxy of the growth of the US economy. The DataStream output is generated through the International Monetary Fund’s financial statistics (IMF)
2. Unemployment - another proxy of the economic growth of the US, also based on IMF numbers
3. Inflation – accounting for the variation in total return through changes in the rent level and is given by the Consumer Price Index (CPI) via the United States Bureau of Labour Statistics
4. Interest rate – a proxy for money availability in the macroeconomic space. The long term interest rate is given by the ten-year US Treasury bill.

2.3 Definition of the model

The thesis will take its offset in a log-linear multiple regression model of the real estate return time series regressed on the mentioned macroeconomic time series given by:

$$\log TBI_t = \beta_1 + \beta_2 \log UNEMP_t + \beta_3 \log CPI_t + \beta_4 \log INT_t + \beta_5 \log GDP_t + \mu_t$$

Where the time series $TBI_t$ is the real estate return index variable, $UNEMP_t$ is the unemployment rate, $CPI_t$ is the inflation rate, $INT_t$ is the long-term interest rate and $GDP_t$ is the growth rate of the United States economy plus a random error term $\mu_t$.

11 All data are available from the appendix 1 page 2-4
12 TBI = Transaction-Based Index. The real estate return term will be described in following pages
2.4 Literature review

The increasing interest in the real estate investment universe in recent years has naturally caught the interest of academicians. Even though the literature on macroeconomic determinants on real estate is limited, comparison of the different available articles shows similar outcome i.e. that understanding of -and management skills within the macro economy is a necessity to obtain good real estate returns.

Case, Goetzmann, and Rouwenhorst (2000) explored returns in global property markets, and found the returns heavily related to fundamental economic variables, while Ling and Naranjo (1997, 1998) identified growth in consumption, real interest rate, the term structure of interest rate, and unexpected inflation as systematic determinants of real estate returns. Hekman (1985) highlights GDP as being the most important influence on return levels, whereas unemployment rate was found not to have any significant impact. The insignificance of employment was backed up by Dobson and Goddard (1992) findings, coinciding with the De Wit and Van Djik (2007) conclusions.

The most extensive research of real estate and the macro economy is in terms of real estate’s hedging capabilities against inflation. Hartzell et al. (1987), Wurtzebach et al. (1991) and Bond and Seiler (1998) proved that real estate provides an inflation hedge across property sectors, and the findings was confirmed by Liu, Hartzell, and Hoesli (1997) as well as Huang and Hudson-Wilson (2007) that found United States real estate market having good hedging abilities. The latter even analyzed the property sectors individually and found that office and residential by far outperformed retail and industrial regarding inflation hedge.

It is however noticeable that Stevenson et al (1997) found no signs of selection ability among investment managers, while there is evidence of superior market timing ability i.e. managers are capable of actively using the macroeconomic environment in order to achieve
superior returns. McIntosh (1989)\textsuperscript{13} equivalently claims that market timing is the key to a successful strategy. Thus the macroeconomic surroundings are evidently important in timing the investments to the property market.

Lastly, Giorgiev, Gupta and Kunkel (2003)\textsuperscript{14} indicated the presence of autocorrelation in direct real estate market returns, which they explain by the appraisal-based valuation method and smoothing problems of direct market indexes. Alternatives to the valuation method in order to avoid the problematic consequences of smoothing will be discussed in chapter 3.

\textsuperscript{13} McIntosh, 1989, page 35-36
\textsuperscript{14} Giorgiev, Gupta and Kunkel, 2003, page 28-33
Chapter 3
Property Valuation and Unsmoothing

In this chapter the complications of property valuation is discussed and methods of correcting for smoothing of real estate returns are investigated. The purpose of the section is to identify the best measure of real estate return to be used in the empirical econometric analysis. The analysis will focus on the individual property valuation; hence the non-listed real estate market is of relevance. However, in order to get a general idea of the market, the different possibilities of property investment will be touched upon briefly.

Real estate investment among professional investors occurs either directly between a buyer and a seller or indirectly via listed or non-listed investment vehicles. Direct investments involve the entire process of acquisition and management of the asset, while indirect investments typically consist of pooled monetary commitments to funds or real estate investment company (REITs). Below table shows an overview of the classic real estate indexes available in the United States.

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Since</th>
<th>Current Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCREIF Property Index</td>
<td>Individual Property</td>
<td>1978</td>
<td>Quarterly</td>
</tr>
<tr>
<td>S&amp;P REIT Composite Index</td>
<td>REITs</td>
<td>1997</td>
<td>Daily</td>
</tr>
<tr>
<td>NAREIT</td>
<td>REITs</td>
<td>1972</td>
<td>Real-Time</td>
</tr>
<tr>
<td>Goldman Sachs Morgan Stanley REIT Index</td>
<td>REITs</td>
<td>1996</td>
<td>Real-Time</td>
</tr>
<tr>
<td>Wilshire Real Estate Indexes</td>
<td>REITs and REOCs</td>
<td>1978</td>
<td>Daily</td>
</tr>
<tr>
<td>Dow Jones REIT Indexes</td>
<td>REITs</td>
<td>1990</td>
<td>Real-Time</td>
</tr>
</tbody>
</table>

Table 3.1 Giorgiev et al. (2003) Benefits of Real Estate Investment, the Journal of Portfolio Management
The value of listed real estate securities equals the day-to-day exchange traded market price of the particular investment company. According to Georgiev et al. (2003) this price does not necessarily mirror the actual value of the underlying asset\(^{15}\), as listed securities to a large extent is influenced by stock market dynamics rather than real estate fundamentals. The uniqueness of private market properties, however, makes valuation much trickier than returns of listed securities. The fundamental problem of real estate is its heterogeneous nature. It is a unique, infrequent and irregularly traded asset between one seller and one buyer – all elements being substantially different from those of the public securities market.

The market value of an individual building is a trade-off of the reservation prices of the buyer and seller of the property. Geltner, Miller et al. (2007)\(^ {16}\) claim the reservation prices of the buyer and seller are overlapping and distributed around a normal distributed point indicating the market value. Hence, the true market value will be an estimation of the different value indications.

Valuation\(^ {17}\) of individual properties can be based on two different approaches:

- Appraisals in property values
- Transaction prices of properties

The first empirical value indicator is appraised value estimates of properties. The appraisals are dispersed cross-sectional around the true market price at a given point in time. The method is based on the existing information on the real estate market values, and is hence the result of a rational behaviour adding the anticipated capital appreciation on the asset on the last known property value, consequently being biased to the previous valuation period.

\(^{15}\) Georgiev, Gupta, and Kunkel, 2003, page 28
\(^{16}\) Geltner, Miller et al., 2006, page 660
\(^{17}\) Valuation refers only to the capital component of real estate return. The income component refers to the rent level of the return and is not biased to property valuation
The absence of a market-based pricing mechanism determines the need for appraisal-based valuation of real estate. The NCREIF National Property Index (NPI) represents valuation-based price the best and an explanation will follow next.

The second and primary indicator of the market value of the direct real estate market is the transaction prices i.e. the trade-off price negotiated by the buyer and seller on the single asset deal. This gives the best empirical indicators of the probability-density distribution of the true market value of the real estate assets. The best represented US index for transaction-based index (TBI) is developed by the Massachusetts Institute of Technology Center of Real Estate (MIT/CRE). This will be discussed further in the coming paragraphs.

### 3.1 NCREIF National Property Index (NPI)

NCREIF (The National Council of Real Estate Investment Fiduciaries) is an association of real estate professionals including investment managers, plan sponsors, academicians, consultants, appraisers etc, and acts as an independent and non-partisan repository of information on real estate performance\(^{18}\).

The quarterly NCREIF National Property Index (NPI) was first produced in 1978, and shows the real estate performance returns submitted from the members weighted by market value, and to a great extend acts as US benchmark for the industry. The index consists of both equity and leveraged properties, however all figures are reported on an unleveraged basis which is also mirrored in the index as a whole. The index includes the four major industry sectors which are returns on apartments, industrial, office and retail properties with all valuations being based on real estate appraisals. The NCREIF NPI universe of properties included in the index is as follows\(^{19}\):

- Existing properties

\(^{18}\) [http://www.ncreif.com/about/](http://www.ncreif.com/about/)

• Only investment-grade, non-agricultural, income-producing properties
• The database is updated quarterly based on transactions performed by participants and data submitted by new NCREIF members
• Sold properties are removed from the database in the quarter the sale takes place, but historical data remains in the database
• Each property’s market value is determined by real estate appraisal methodology, consistently applied

The return reported in the database is the total rate of return, which is calculated by adding the income return to the capital appreciation (depreciation) return on a quarterly basis. Unfortunately for non-members the valuation-based index is only available on an aggregate total return basis; hence it is not possible for third parties to obtain returns on both income component and capital appreciation component. Therefore the data is not useful for econometric analysis, as it makes it impossible to unsMOOTH the capital value. I will return to this in later paragraphs.

3.2 MIT Transaction Based Index (TBI)

MIT (Massachusetts Institute of Technology) Centre for Real Estate was founded in 1983 and includes a leading academic research and publication entity in the real estate industry and was moreover founder of the first Master of Science in Real Estate Development. MIT reported their index from 1984 across all sectors, whereas the individual sector index has only been calculated from 1994 and forward.

The data laboratory of MIT has developed a transaction-based index (TBI) on US Institutional commercial property investment performance, based on properties sold from the NCREIF Index database. Hence it is a direct statistical outcome of the NCREIF data. Where the valuation-based returns from NCREIF are based on appraisal estimates, the TBI mirrors the actual transaction prices of the properties, and hereby improves the measuring
of the movements of the industry returns, but in the same time also increasing the volatility of the real estate returns.

The TBI is developed through a hedonic pricing model by Fisher, Geltner and Pollakowski (2006), decomposing the NCREIF data and using the latest appraised value of each transacting property as a hedonic variable\textsuperscript{20} in order to avoid index smoothing and lagging biases of the index. Hence, the model uses dummy coefficients to represent the time difference between appraisals and the transaction prices.

The unsmoothed returns of the hedonic TBI model, makes the model suitable for further econometric analysis. However even though the model removes the issues of smoothing and temporal lag biases, it can contain other estimation errors. The critical issues of statistical errors will be discussed shortly.

### 3.3 Comparison of NPI and TBI

The exhibit below compares the MIT transaction-based index TBI with the NCREIF Property Index NPI and shows some of the characteristics of each valuation method. The smoothing of the appraisals-based NPI appears to smooth out the time series, making it less volatile than the transaction-based index. Also TBI appears to lead the trend curve of NPI in turning points of the historical cycle of the real estate industry and macro-economic surroundings. This can also be explained by the temporal lag bias of the NPI making it less affected my major events.

Fisher, Geltner and Pollakowski (2006) also highlight less autocorrelation and less seasonality in the transaction-based index compared to appraisal-based indexes. All of the above will be further analysed in the econometric study from chapter 4 an onwards.

\textsuperscript{20} Fisher, Geltner, Pollakowski, 2006, page 1
The figure also shows a demand side measure of the TBI referred to as constant liquidity index. The demand side index collapses price and trading volume into the same metric, showing the constant time on the market or constant turnover ratio of trading volume subject to the percentage change in property price\textsuperscript{21}. The characteristics of the constant liquidity index will not be analysed further in this paper.

### 3.4 Noise and appraisal error in real estate valuation

The two valuation approaches described are both indicators of the market value of properties; however they both suffer from certain statistical noise. The difference between the actual transaction price and the unobservable *true* market price are considered transaction price error (random error). The noise is unbiased as the transaction price randomly differs from the true market value. The appraised values are as stated biased as they are lagged in time from previous period’s values. This is referred to as temporal lag bias\textsuperscript{22}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.2.png}
\caption{Jim Clayton, Director of Research, Pension Real Estate Association (PREA) 2008}
\end{figure}

\textsuperscript{21} Fisher, Geltner, Pollakowski, 2006, page 5-6
\textsuperscript{22} Geltner, Miller et al, 2006 page 659-661
Hence, there are two major concerns in terms of real estate valuation errors:

- Transaction price error (random error)
- Temporal lag bias (biased error)

Unfortunately a natural trade-off makes it difficult to reduce either of the errors without increasing the other error. Geltner, Miller, Eicholtz et al. (2006) have developed a theoretical noise-lag trade-off frontier showed in following figure\(^\text{23}\), where the utility of the valuation method is optimized by reducing both the random error and the temporal lag bias, i.e. highlighting the issue of which valuation method is optimal to use working with real estate values and returns.

The vertical axis represents temporal lag biases, which is minimised the farther along the axis one goes. The theoretical perfection is zero lag bias at point 0. The horizontal axis represents greater preciseness in the value estimates the farther along the axis one goes. The disaggregate isoquant \(T_{\text{Dis}}\) represents the individual property appraisals, whereas the aggregate isoquant \(T_{\text{Agg}}\) is relevant for the macro-level index construction that is of interest in this paper. In order to maximise the utility of the function, \(U\) ought to move towards the upper-right corner reducing the errors in the estimation, however still be tangent to the individual valuation \(T_{\text{Dis}}\). The theoretically optimal aggregate valuation is utility function

\(^{23}\) Geltner, Miller et al., 2006, page 663
U₁, which according to Geltner et.al (2006) is best represented by a regression-based transaction price index equivalent to the MIT transaction-based index.

### 3.4.1 Effects of the errors

There are numerous effects of the random errors attached to the two index types. In the coming paragraph the effects will shortly be highlighted to show the difficulty of the trade-off between the errors.

**Transaction price errors – pure random error effects**

- Spurious error terms increase volatility
- Reduces any positive first-order autocorrelation
- Reduces apparent cross-correlation with exogenous series

The return of the observable transaction-price based capital appreciation deviates from the true return of the unobservable capital appreciation. The expected value of the random error is as always zero, however, the pure random errors makes the empirical return more volatile. Spurious error terms cause the observed values to vary across the true unobservable return increasing the standard deviation over time. The noise also reduces any positive first-order autocorrelation of the true returns. Lastly, the increased volatility also reduces any cross-correlation between the observable returns and any exogenous series.

**Appraisal based indices - temporal lag bias effects**

- Lagged time series in the form of a moving average
- Incorrectly reduces the volatility
- The return is biased towards the preceding periodic trend
- Increase the (positive) autocorrelation
- Reduces the systematic risk of the real estate return

---

25 The potential correlation between consecutive true returns is reduces due to the random noise
26 Geltner, Miller et al, 2006, page 667-669
The lagging of the time series in the appraisal-based valuation imply a moving average of the returns, and incorrectly reduces the volatility. Further the moving average process will bias the direction of the return, depending on the preceding period’s trend i.e. in turning points; the observable returns will be biased toward the previous quarter and perhaps adjust less than anticipated. Next, the bias of the moving average lagged series will to a bigger degree affect the positive autocorrelation in the empirical returns, than that of the unobservable true returns.

The undervalued volatility additionally implies a reduced systematic market risk of real estate returns compared to a non-lagged risk measurement. This appraisal smoothing effect causes an underestimating of the risk associated to the real estate industry, which potentially could cause an over-allocation by investors to the real estate segment.

The comparison below summarises the pure effect of the statistical errors. The true value is the unobservable true market value. The noise of the random error increase the volatility, whereas the lag-effect smoothes the value, making it less volatile. The noise-lag trade off is represented by the purple appraisal-based valuation line, including both types of errors and hereby concealing the actual effect of the two statistical errors.

Figure 3.4: Geltner, Miller et al. (2007). Lecture notes, slide 50
3.5 The methodology of moving averages

This paragraph will concisely describe the theory of a moving average. The actual methodology of unsmoothing real estate returns are a reverse technique of the moving average approach. Thus, the topic will only be touched upon briefly for information on the common smoothing process. A moving average is smoothing of a time series, used when the aim is to express a general level of a series. There are a variety of averaging techniques that will not be explained further in this paper including the simple moving average, centred moving average, double moving average and weighted moving averages. The technique relevant in the real estate index unsmoothing is reversal of a simple exponential smoothing process.

The mathematical simple exponential smoothing model was conceived by Macaulay (1931) and later developed by other academicians for forecasting purposes. The basic implication of such a model is an expectation of exponentially declining effects from observation over time, i.e.:

\[ MA_{t+1} = (\alpha)MA_t + (1-\alpha)x_t \]

Where \( MA_{t+1} \) is the moving average prediction one period ahead, \( x_t \) is the average value of the observations at time \( t \). The fraction of the latest observation is the smoothing constant \( \alpha \). The critical point of the model is the value of the smoothing constant \( \alpha \) that ranges between \([0; 1]\). A smoothing constant close to 1 gives more weight to the most recent observation and less to distant observations, while the opposite is given by a smoothing constant close to 0.

One can discuss the need of correction for trends, seasonality and cycles in real estate returns. The real estate market is proven to be comparable to the general economic development in the given area, i.e. some cyclical trend would be expected. However, given

---

27 Yaffee et. al, 2000, page 18-23
28 Yaffee et. al, 2000, page 23
that the data are only provided on a quarterly basis, the question is to what degree the irregularities are observable and to what extent they are present. The features are neither taken into consideration in the unsmoothing theories of real estate returns; hence I will ignore the issue for now. In the following three unsmoothing processes of the capital appreciation is discussed, where the implications of a reverse technique of such a simple exponential moving average model is incorporated. Note that unsmoothing is only relevant for the capital component of the total return, whereas the income component given by the property’s rent level is irrelevant at this stage.

3.6 Methods of unsmoothing appraisal-based returns

During the last couple of decades a number of unsmoothing models have been developed in order to avoid the lag bias problem of the appraisal-based returns. Following are three main approaches to adjust the returns for smoothing effects:

1. Zero-autocorrelation technique
2. Mechanical de-lagging (mathematical approach)
3. Transaction based regression modelling (econometric approach)

3.6.1 Zero-autocorrelation technique

The first approach is a basic unsmoothing model still widely used in academic research. The rationale of the model is simply that if real estate returns were liquid and efficient they would be uncorrelated. The technique is statistically to remove the autocorrelation from the appraisal-based real estate returns via the residuals.

Assuming quarterly returns, Geltner and Miller et al. (2007) recommend a first- and fourth-order autoregression on the observed returns. The residuals are then adjusted by a

---

29 Geltner, Miller et al, 2007., page 681
30 Geltner, Miller et al, 2007., lecture notes, slides 68-70
31 Geltner, Miller et al., 2007, lecture notes, slides 68-69
constant factor to produce a reasonable mean and volatility. The outcome of such a regression is shown from the figure below, based on the NPI quarterly appreciation returns in the period from 1979-2005\textsuperscript{32}. Through the regression both lag bias and volatility have been adjusted, which increase the standard deviation of the unsmoothed appraisal returns. The weakness of the model is the manually chosen constant factor that entirely determines the unsmoothing process, which makes model specification bias almost inevitable.

![Graph showing NPI appreciation and unsmoothed capital value index from 1978 to 2005.](image)

Figure 3.5: Own production based on Geltner, Miller et al 2007, Appendix 25B.

### 3.6.2 Mechanical de-lagging (mathematical approach)

The mechanical de-lagging approach is used to adjust the appraisal lag bias to retrieve a contemporaneous transaction price value via reverse-engineering, i.e. going backwards in the appraisal process by reversing the lagging of the exponential weighted moving average model.

\textsuperscript{32} See the zero-autocorrelation calculations from disclosed CD-Rom
The optimal mechanical unsmoothing model is according to Geltner, Miller et.al (2007) and Marcato and Key (2007) a first-order autoregressive model as follows\textsuperscript{33}:

\[ V_t^* = \omega_0 \bar{V}_t + (1 - \omega_0)V_{t-1}^* \]

Where the transaction price can be approximated with the lag weights being the same in any two adjacent lag: \( \omega_{L+1}, \omega_L = \rho < 1 \) and in which \( V_t^* \) is the current appraisals, \( V_{t-1}^* \) is the previous appraisal and \( \bar{V} \) is the contemporaneous transaction price evidence. Assuming log levels the contemporaneous transaction-based return can then be derived from following\textsuperscript{34}:

\[
\begin{align*}
    r_t^* &= \omega_0 r_t + (1 - \omega_0) r_{t-1}^* \\
    r_t &= \frac{r_t^* - (1 - \omega_0) r_{t-1}^*}{\omega_0}
\end{align*}
\]

Again, \( r_t^* \) is the appraisal-based return, \( r_{t-1}^* \) is the appraisal-based return from the previous period and \( r_t \) is the contemporaneous transaction-based return. The inverse appraisal-based return equals the contemporaneous transaction-based return \( r_t \). The weight \( \omega_0 \) is given by \( \omega_0 = 1/(\bar{L} + 1) \), where \( \bar{L} \) is the average number of periods of lag in the appraisals com-

\textsuperscript{33} Geltner, Miller et al, 2007, page 682
\textsuperscript{34} Geltner, Miller et al, 2007, page 682
pared to the current transaction price\textsuperscript{35}.

![Graph comparing appraisal based NPI and transaction based TBI](image)

The graph above compares the appraisal based NPI and the transaction based TBI and gives an intuition of the number of time lags between the time series. For quarterly data as that of the NPI the number of lags are theoretically anticipated to be about four periods i.e. $L = 4$ leading to a weight of $\omega_0 = 1 / (4+1) = 0.2$, the graph however shows differences in lags depending of the time period between two to four lags. Yet, in order to be consistent the following example will anticipate a lag of four periods in the model.

Given the return formula and assuming four lags in the weight $w_0$, the contemporaneous transaction-based returns, based on the logs of the quarterly NPI capital appraisal returns, are as follows:

\textsuperscript{35} Geltner, Miller et al, 2007, page 682
The impact of the mechanical de-lagging process is as expected an increased volatility of the returns of the contemporaneous transaction-based returns given by $r_t$ calculated from the smoothed appraised returns $r_t^*$. Be aware of the risk of the properties not being reappraised in each period, in which case the unsmoothed returns suffer from a missing valuation observation. This will also inflict a lag-effect on the returns. Unfortunately the problem is not apparent from the NCREIF data.

In addition to the first-order autoregressive reverse filter mentioned above, Marcato and Key (2007) highlights three other unsmoothing techniques\textsuperscript{36}, namely (i) a more generalized second-order autoregressive filter model by Geltner (1993), taking yet another autoregressive process into the model, which may be useful in case of frequent appraisals. However, the AR(1) also captures previous periods lags indirectly which devalues the impact of the AR(2) approach. (ii) Fisher, Geltner and Webb (1994) developed a full information value index, where the volatility of the residuals are used as a weight on the first-order autoregressive specification model and finally (iii) Chaplin (1997) assumed the unsmoothing parameter i.e. the weight of the lagged returns should differentiate depending

\textsuperscript{36} Marcato and Key, 2007, page 89-91
on the upward or downward trend of the underlying market. Thus, the unsmoothing once again is based on a first-order autoregressive model, but now with varying unsmoothing parameters. The unsmoothed total return is then derived by adding the unsmoothed capital appreciation return to the income return given by the rent level of the property.

Other procedures as e.g. time-varying methods and a less sophisticated simple one-step model for annual frequency data also exist\textsuperscript{37}, but these are of less relevance to the NPI unsmoothing process, and will not be highlighted here.

\textbf{3.6.3 Regression modelling (econometric approach)}

Another approach to estimate the real estate valuation and return is the transaction-based index also used by MIT to produce the TBI index. This statistically conservative approach uses the actual contemporaneous transaction price values in a regression model specified to avoid the appraisal lag biases.

The two main approaches for constructing such a real estate index are a hedonic model approach and a repeat-sales approach. Last mentioned is a dummy-variable model only including properties transacted at least twice during the sample period, and using dummy-variables to identify the changes in the market over time\textsuperscript{38}.

The far most common transaction based index is the hedonic regression model based on all transactions of the period. A substantial number of qualitative and quantitative property-specific cross-sectional explanatory variables are used in the regression model, which describe the characteristics that affect property values\textsuperscript{39}. The hedonic log-function typically contains 60 or more explanatory variables varying from number of rooms, quality of isolation, property interior, location etc.\textsuperscript{40}. The strength of the hedonic approach is that it is founded on actual transactions, however this unbiased estimation anticipates that all

\textsuperscript{37} Geltner, Miller et al, 2007, page 683 and Marcato and Key, 2007, page 89-91
\textsuperscript{38} MIT Center for Real Estate Transaction Based Index; \url{http://web.mit.edu/cre/research/credl/tbi.html}
\textsuperscript{39} Geltner, Miller et al, 2007, page 683 and Marcato and Key, 2007, page 684
\textsuperscript{40} See appendix 2 page 12 for list of potential variables from Kagie and Wezel, 2008
explanatory variables have been taken into consideration, which create a risk of an *omitted variable bias*, consequently misspecification of the model is the main vulnerability of the hedonic regression model\(^\text{41}\).

The transaction-based index derived by MIT from the NCREIF data is exactly such a hedonic index. The index is based on an assessed-value approach developed by Clapp and Giacotto that reflects the property characteristics relevant for determining the property values as a composite hedonic variable\(^\text{42}\). The outcome of the regression is a correction of the lag bias between the appraisals and the current transaction prices. The hedonic regression model used by MIT is as follows\(^\text{43}\):

\[
P_t = \sum_j a_j x_{ijt} + \sum_i \beta_i z_i + \epsilon_t
\]

- \(P_t\) are logs of transaction prices (property i, period t)
- \(X_{ijt}\) is a vector of j hedonic variables
- \(Z_i\) is a time dummy variable, where 1 if sale i occurred in time t, 0 if otherwise

As mentioned previously there is a natural trade-off of estimation errors between the valuation-based and transaction-based indices. In practice real estate investors tend to focus on the NCREIF valuation-based index NPI, due to its information depth and easy accessibility. Academicians, however, by far prefer the transaction-based approach as that of TBI developed by MIT because of its consideration of bias correction and its better estimation of the true market values. Thus, the following econometric analysis of the macroeconomic determinants of the US real estate return will be based on the unsmoothed returns of TBI.

### 3.7 Introduction to econometric analysis

\(^{41}\) Andersen and Hjortshøj, 2008, page 10  
\(^{42}\) MIT Center for Real Estate Transaction Based Index; [http://web.mit.edu/cre/research/credl/tbi.html](http://web.mit.edu/cre/research/credl/tbi.html)  
\(^{43}\) Geltner, Miller et al, 2007, Lecture Notes, Slide 82
By now we have identified the relevant return variable and discussed the valuation issues of real estate. Moving forth, the actual empirical analysis will be performed based on the log-linear regression model defined previously including the transaction-based real estate return index as dependent variable. The model was given by:

$$\log TBI_t = \beta_1 + \beta_2 \log UNEMP_t + \beta_3 \log CPI_t + \beta_4 \log INT_t + \beta_5 \log GDP_t + \mu_t$$

The analysis will be performed econometrically through regression analyses; hence the model must concur to the assumptions of the classical linear regression model. Before the assumptions are to be tested in chapter 7, the econometric terms of stationarity and co-integration will be investigated in chapter 4 and 5.
Chapter 4
Stationarity of the Time Series

Prior of using the macroeconomic variables in a combined empirical analysis it is essential to understand the properties of the individual time series variable. In order to study the behavior of the time series correctly the data needs to be stationary, meaning its mean, variance and auto-covariance at various lags must be time invariant\(^44\) i.e. there is no reversion to the mean. A stationary variable is integrated of order zero. If it is concluded the two series are integrated of the same order, tests concerning co-integration can be performed via a unit root test on the residuals from the regression model. In the case of non-stationary variable the problem of spurious regression may arise in the multivariate regression model\(^45\).

The test of stationary time series will be performed both graphically and numerically by the time domain approach modeling the lagged relationship directly on the time series. The restrictions that are to be fulfilled are the following all to be independent of time\(^46\):

\[
E(Y_t) = \mu \\
\text{var}(Y_t) = \sigma^2_Y \\
\text{cov}(Y_t, Y_{t-k}) = \gamma_k
\]

\(^{44}\) Gujarati 2003, page 798
\(^{45}\) Koop, 2008, page 177
\(^{46}\) La Cour, 2008, Introduction to time series analysis, slide 6
The violation of the stationary series is if $Y_t$ contains a stochastic trend or a deterministic trend. Stationary time series is said to be integrated of order zero. Non-stationary time series, however, can be made stationary by taking its difference $d$ times i.e. it is integrated of order $d$. In case of several time series are integrated of the same order $d$, co-integration might be present among the variables. I will return to co-integration analysis in chapter 6.

In practice testing for stationarity is performed in SAS for each regression model variable as the dependent variable. Depending on the line plot of the dependent variable against time, the variables are a lagged version of them and potentially time $t$. The effects of these can be seen from the following and appendices on stationarity.

### 4.1 Stationarity tests for TBI

The log of the transaction based index of the US real estate returns from 1984:1-2008:1 will as mentioned act as the dependent variable of the multiple regression model in the empirical analysis. In order to assure stationarity in the variables graphical and numerical tests are performed to test the properties of the time series.

1) **Graphical stationarity test:**

Firstly a line plot of the log value of the TBI change against time from 1984:1-2008:1 will show the historical behavior of the time series.

![Figure 4.1 Changes in log values of TBI – see appendix 1 page 5](image)
The immediate interpretation of the graph indicates a non-stationary process with non-zero mean and a stochastic trend in the TBI variable i.e. the TBI time series will not be stationary, but follow a random walk with a trend. The graphical outlook of the variables will be examined further in the following numerical tests.

2) Model estimation and numerical stationarity tests:

Secondly the anticipated model is estimated and a number of numerical stationarity tests are performed in mentioned order:

1. Durbin-Watson d-statistic
2. The Breusch-Godfrey test (Lagrange Multiplier test)
3. ACF / PACF correlograms
4. The (augmented) Dickey-Fuller Unit Root Tests

The numerical test is a regression of the delta TBI as dependent variable against the lagged TBI and time as explanatory variable, all values being in logs. Assuming the graphical test indicates a pure random walk with a stochastic trend of the time series, the function is the following, where \( \Delta \) denotes the first difference operator that is \( \Delta TBI_i = TBI_i - TBI_{i-1} \)

\[
\Delta TBI_i = \beta_1 + \beta_2 TBI_{i-1} + \beta_3 t + \mu_i
\]

The estimated model is then

\[
\Delta \hat{TBI}_i = 0.0139 - 0.0228 \hat{TBI}_{i-1} + 4.273E^{-11}t + \mu_i
\]

<table>
<thead>
<tr>
<th>SE</th>
<th>( t )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.0225)</td>
<td>(0.62)</td>
<td>(0.5391)</td>
</tr>
<tr>
<td>(0.0211)</td>
<td>(-1.08)</td>
<td>(0.2819)</td>
</tr>
<tr>
<td>(2.605E-11)</td>
<td>(1.64)</td>
<td>(0.1042)</td>
</tr>
</tbody>
</table>
4.1.1 Durbin-Watson d-statistic

Firstly, according to Granger and Newbold, the *Durbin-Watson d-statistic* indicates the accuracy of the regression, the output is found from the Yule-Walker Estimates of the single time series regression model\(^47\). The test statistic is as follows\(^48\):

\[
d = \frac{\sum_{t=2}^{n} (\hat{\mu}_t - \hat{\mu}_{t-1})^2}{\sum_{t=1}^{n} \hat{\mu}_t^2}
\]

With the null-hypothesis:

\[H_0: \rho = 0 \text{ (no (positive) autocorrelation)}\]
\[H_1: \rho > 0\]

Where \(\hat{\mu}_t\) is the residual associated with the observation at time \(t\) and it is assumed that:

- An intercept term is included in the regression model
- Non-stochastic explanatory variables
- First-order autoregressive disturbance terms \(\mu_t\)
- Normally distributed error term \(\mu_t\)

The Yule-Walker estimate of the regression model is:

<table>
<thead>
<tr>
<th>Yule-Walker Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
</tr>
<tr>
<td>MSE</td>
</tr>
<tr>
<td>SBC</td>
</tr>
<tr>
<td>Regress R-Square</td>
</tr>
<tr>
<td>Durbin-Watson</td>
</tr>
</tbody>
</table>

Table 4.1 TBI Durbin-Watson test\(^49\)

\(^{47}\) Gujarati page 806-807

\(^{48}\) Gujarati page 467: Durbin-Watson d statistic

\(^{49}\) Appendix 3 page 15
As the $d$ statistic of 1.9823 > $R^2$ of 0.0588 it implies no sign of a spurious regression and the $d$-statistic also lies within the 5% Durbin-Watson critical value of [1.645; 2.355] with one explanatory variable and 94 number of observations indicating no significant autocorrelation\(^{50}\).

### 4.1.2 The Breusch-Godfrey Series Correlation test (LM test)

The regression model also performs the Lagrange Multiplier test (LM test) based on Breusch and Godfrey's test of autocorrelation. The test is very general and allows for non-stochastic regressors, higher-order autoregressive schemes and simple or higher-order moving averages of white noise errors\(^{51}\). Oppositely the Durbin-Watson test allowed no lagged values of the regressand among the regressors.

The null-hypothesis of the test is no serial correlation in the error terms of any order\(^{52}\):

\[
H_0: \rho_1 = \rho_2 = \ldots = \rho_p = 0
\]

\[
H_1: \text{Not } H_0
\]

SAS automatically calculates the test with four lagged values of the residuals i.e. it accounts for autoregressive schemes of order four, which ought to be sufficient amount. The test statistic is given by $(n-p)*R^2$ obtained from the auxiliary regression of the estimated residuals regressed on the explanatory variables of the original model. The test follows the chi-square distribution with $\rho$ degrees of freedom, which in this case is four degrees of freedom\(^{53}\).

<table>
<thead>
<tr>
<th>Godfrey's Serial Correlation Test</th>
<th>Alternative</th>
<th>LM</th>
<th>Pr &gt; LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.0078</td>
<td>0.9295</td>
<td></td>
</tr>
</tbody>
</table>

\(^{50}\) Gujarati page 970, Appendix D, table D.5s: Durbin-Watson d Statistic

\(^{51}\) Gujarati page 473-474: The Breusch-Godfrey Test

\(^{52}\) Gujarati page 473: The Breusch-Godfrey Test

\(^{53}\) Gujarati page 473-474: The Breusch-Godfrey Test
Table 4.2 Breush-Godfrey Serial Correlation Test\(^{54}\)

<table>
<thead>
<tr>
<th>Alternative</th>
<th>LM</th>
<th>Pr &gt; LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(2)</td>
<td>0.1076</td>
<td>0.9476</td>
</tr>
<tr>
<td>AR(3)</td>
<td>0.1104</td>
<td>0.9906</td>
</tr>
<tr>
<td>AR(4)</td>
<td>3.8421</td>
<td>0.4278</td>
</tr>
</tbody>
</table>

The SAS test output is to be interpreted the following way: The test statistic, here given by the LM column, shows the relevant statistic of the test given its appropriate distribution. The Pr column to the far right, here given by Pr>LM, indicates the P-value at a 5% significance level.

The hypothesis of no serial correlation in the TBI cannot be rejected, as they are all within the 5% critical value of 9.48773 in the chi-square distribution. Hence it backs up the Durbin-Watson test results and indicates no serial correlation of any order.

### 4.1.3 Correlograms of the ACF and PACF

Correlograms show the correlation between a variable and a lag of itself, and is a useful way to understand the properties a time series and a simple test of stationarity. The correlogram from SAS shows the autocorrelation function (ACF) and the partial autocorrelation function (PACF) which imply the significance of the autocorrelation within the model. If the autocorrelations at various lags hover around zero, the time series follows a stationary process.

The ACF at lag k denoted by \(\rho_k\) is defined as\(^{55}\):

---

\(^{54}\) See appendix 3 page 14
\(^{55}\) Gujarati page 808, Autocorrelation Function
\[ \rho_k = \frac{\gamma_k}{\gamma_0} \]

\[ \rho_k = \frac{\text{covariance at lag } k}{\text{variance}} \]

The lag length of the time series has been settled at 24, which corresponds to approximately one fourth of the time period of the series. The choice matches the rule of thumb mentioned in Gujarati, 2003\(^{56}\).

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.076908</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.074556</td>
<td>0.96941</td>
<td>1.01015</td>
</tr>
<tr>
<td>2</td>
<td>0.072082</td>
<td>0.93724</td>
<td>0.171414</td>
</tr>
<tr>
<td>3</td>
<td>0.069535</td>
<td>0.90413</td>
<td>0.217508</td>
</tr>
<tr>
<td>4</td>
<td>0.066819</td>
<td>0.86881</td>
<td>0.252967</td>
</tr>
<tr>
<td>23</td>
<td>0.023845</td>
<td>0.31004</td>
<td>0.452960</td>
</tr>
<tr>
<td>24</td>
<td>0.021987</td>
<td>0.28589</td>
<td>0.455121</td>
</tr>
</tbody>
</table>

Table 4.3 TBI ACF and PACF Correlograms

The above ACF and PACF correlograms\(^{57}\) show the behavior of the functions in the first four lags and the last two of the lag length (last three for PACF). Given the decaying behavior of the ACF it can be concluded that the TBI time series is non-stationary that is there may be non-stationarity in mean or variance or both.

---

\(56\) Gujarati page 812: Choice of lag length in autocorrelation functions

\(57\) See appendix 3 page 16-17
The solution to the non-stationarity problem is to transform it into a stationary time series by taking its first difference. Below is shown an equivalent table of the I(1) function of the TBI series, and from that it becomes clear that the problem of non-stationarity is no longer present, in spite of the borderline spike in lag four, thus the TBI is stationary integrated of order 1.

### 4.1.4 Dickey-Fuller Unit Root Test (DF-test)

Lastly, looking towards the \textit{(augmented) Dickey-Fuller Unit Root Tests} (DF-test) corrected for any correlation and with the null-hypothesis of non-stationarity $H_0 = \delta = 0$ (non-stationarity) the tau value tests stationarity for time series with zero mean, non-zero mean and non-zero trend. Above it was concluded that the time series for TBI had a non-zero mean and a trend, accordingly both will be assessed at various lags. If testing the time series assuming it followed an I(0) the non-stationarity test of Dickey-Fuller would not be rejected, however the following table shows the DF-test when the series has been integrated once.

#### Table 4.4 TBI ACF and PACF Correlograms differenced once

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
<th>Autocorrelations</th>
<th>Partial Autocorrelations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00025416</td>
<td>1.00000</td>
<td><img src="https://example.com/autocorr.png" alt="Autocorrelations" /></td>
<td><img src="https://example.com/partial_autocorr.png" alt="Partial Autocorrelations" /></td>
</tr>
<tr>
<td>1</td>
<td>0.06011E-6</td>
<td>0.03175</td>
<td><img src="https://example.com/partial_autocorr.png" alt="Partial Autocorrelations" /></td>
<td><img src="https://example.com/autocorr.png" alt="Autocorrelations" /></td>
</tr>
<tr>
<td>2</td>
<td>1.3092E-6</td>
<td>0.00515</td>
<td><img src="https://example.com/autocorr.png" alt="Autocorrelations" /></td>
<td><img src="https://example.com/partial_autocorr.png" alt="Partial Autocorrelations" /></td>
</tr>
<tr>
<td>3</td>
<td>0.0001033</td>
<td>0.04084</td>
<td><img src="https://example.com/partial_autocorr.png" alt="Partial Autocorrelations" /></td>
<td><img src="https://example.com/autocorr.png" alt="Autocorrelations" /></td>
</tr>
<tr>
<td>4</td>
<td>0.00005762</td>
<td>0.22689</td>
<td><img src="https://example.com/partial_autocorr.png" alt="Partial Autocorrelations" /></td>
<td><img src="https://example.com/autocorr.png" alt="Autocorrelations" /></td>
</tr>
<tr>
<td>23</td>
<td>-5.2001E-6</td>
<td>-0.2046</td>
<td><img src="https://example.com/autocorr.png" alt="Autocorrelations" /></td>
<td><img src="https://example.com/partial_autocorr.png" alt="Partial Autocorrelations" /></td>
</tr>
<tr>
<td>24</td>
<td>-9.1869E-6</td>
<td>-0.3615</td>
<td><img src="https://example.com/partial_autocorr.png" alt="Partial Autocorrelations" /></td>
<td><img src="https://example.com/autocorr.png" alt="Autocorrelations" /></td>
</tr>
</tbody>
</table>

58 See appendix 3 page 18-19
59 The Dickey-Fuller test assumes uncorrelated error terms. The augmented Dickey-Fuller test is corrected for any such correlation
Looking at the single mean and trend for the I(1) series, the tau statistics is clearly significantly different from zero at all lags, meaning the null-hypothesis of non-stationarity easily can be rejected. Hence the TBI time series variable contains a unit root and becomes stationary following an I(1) process.

In conclusion there seems to be contradictive results of the autocorrelation tests. The Durbin-Watson d statistic and the the Breusch-Godfrey test found no signs of autocorrelation, whereas the ACF and PACF correlograms did. However when differenced once to an I(1) process the problem of autocorrelation disappears and the TBI time series becomes stationary.

### 4.2 Stationarity among the explanatory macro variables

Looking towards the explanatory variables of the original multiple regression model, the stationarity test of the macroeconomic variables are all accepted following a process

---

60 See appendix 3 page 19
61 See appendices 3 page 20-43 for test results of the explanatory variables
integrated of order 1. The test results are available from appendix 3. There are two important properties to recognize from the results, first the fact that all variables are integrated of the same order, namely they follow an I(1) process is crucial, as the variables in a regression model must be integrated of the same order to obtain feasible estimates. Second, the unit root properties of the time series endorse the chance of co-integration between some or all the variables. The issue of co-integration and its properties will be tested in the following chapter 5.
Chapter 5
Co-integration among Unit Root Time Series

From the unit root analysis of the variables in the section of stationarity tests, it was proved that all variables in the model are I(1) i.e. they all contain unit roots. However there may be a linear combination between the non-stationary time series that is stationary namely a long term equilibrium between the variables that cancels out the stochastic trends of the series. The basic theoretical intuition of co-integration is as follows:

\[ Y_t = \alpha + \beta X_t + \epsilon_t \]
\[ \epsilon_t = Y_t - \alpha - \beta X_t \]

If the two variables \(X_t\) and \(Y_t\) both contain non-stationary stochastic trends, an equivalent behavior would be expected from the linear relationship between them. However it is possible that the unit roots cancel out each other making the error term of the model stationary. Co-integration thus provides some good economic intuition that there is an equilibrium relationship between the variables or that the variables in question trend together.

Working with macroeconomic variables I would expect some relationship among some or all of the variables. Given the stationarity tests in the previous paragraph it is given that all variables are I(1) i.e. they contain a stochastic trend, hence the fundamentals for potential

---

62 Gujarati, 2003 page 822
63 Koop, 2008 page 217-218
64 Koop, 2008 page 218
co-integration between some of the variables are present as expected. In the following I will shortly describe which co-integration tests that will be performed and the expected outcome of such co-integration test.

5.1 Expectations and performance of co-integration tests

Given the properties of the variables it is expected that the co-integration tests will show some relationship between at least some of the variables. It is however important to notice that a direct relationship might not be applicable even in case of high correlation between them - but instead a number of factors cause an indirect relationship between the variables. For instance high correlation between mortality and alcohol might not result in a direct co-integration relationship, but can be caused by a third-factor influence as e.g. smoking.

The co-integration tests performed in this paper are:

1. Co-integrating Regression Durbin-Watson Test
2. Augmented Engle-Granger Test

In addition to above two mentioned co-integration tests, there exist a number of other more sophisticated instruments for co-integration testing\(^{65}\), however these will not be included in this paper. The co-integration tests have been performed (i) for the regression model as a whole, (ii) for all the explanatory variables in one and (iii) in pairs of all variables and can be found in appendix 4. From appendix 4 can also be seen that testing the overall multivariate model does not result in a co-integration relationship\(^{66}\), therefore instead a model consisting of all the explanatory variables will be tested next.

---

\(^{65}\) Koop, 2008, page 221

\(^{66}\) See appendix 4 page 45-46
5.1.1 Co-integrating Regression Durbin-Watson (CRDW) Test

The Co-integrating Regression Durbin-Watson Test was provided by Sargan and Bhargava in 1983\(^67\). The test takes its basis in the Durbin-Watson d-statistic which tests the null-hypothesis of \( d = 2 \). However, given a unit root the estimated \( p \) will be approximately 1 implying \( d = 2(1-p) \rightarrow d = 2(1-1) = 0 \). In this test the null-hypothesis will therefore be \( H_0: d = 0 \).\(^68\)

The model to be tested is the linear relationship between all the macroeconomic explanatory variables. The regression model has GDP\(_t\) as the regressand and the remaining explanatory variables as regressors. The regression model consists of the log value of US GDP\(_t\) regressed on the log value of the explanatory variables in the period and an error term in the period. The OLS estimates are:

\[
\begin{align*}
\log GDP_t = & \beta_1 + \beta_2 \log UNEMP_t + \beta_3 \log CPI_t + \beta_4 \log INT_t + \mu_t \\
\log \hat{GDP}_t = & 9.285 - 0.3994 \log \hat{UNEMP}_t + 0.1384 \log \hat{CPI}_t - 1.1471 \log \hat{INT}_t,
\end{align*}
\]

\[
t = (24.80) \quad (-4.13) \quad (3.18) \quad (-23.24)
\]

<table>
<thead>
<tr>
<th>Ordinary Least Squares Estimates</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>0.35694674</td>
</tr>
<tr>
<td>DFE</td>
<td>95</td>
</tr>
<tr>
<td>MSE</td>
<td>0.00376</td>
</tr>
<tr>
<td>Root MSE</td>
<td>0.06130</td>
</tr>
<tr>
<td>SBC</td>
<td>-257.57326</td>
</tr>
<tr>
<td>AIC</td>
<td>-267.95374</td>
</tr>
<tr>
<td>Regress R-Square</td>
<td>0.8678</td>
</tr>
<tr>
<td>Total R-Square</td>
<td>0.8678</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>0.5947</td>
</tr>
</tbody>
</table>

Table 5.1 Co-integration test among the explanatory variables\(^69\)

\(^{67}\) Gujarati, 2003, page 824
\(^{68}\) Gujarati, 2003, page 824
\(^{69}\) See appendix 4 page 47
As can be seen from the model the three macroeconomic variables explain an impressive 86% of GDP in the period from 1984:1-2008:2, meaning interest rate, inflation and unemployment grasp quite a bit of the log change in GDP. The model’s estimates act as economically anticipated namely that GDP is negatively influenced by unemployment and interest rate, whereas inflation increases the value of the economic growth. In addition the relationship among the variables is confirmed by the significant estimates of the model.

According to Gujarati the critical values of the CRDW test on 1, 5, and 10 percent levels on the null-hypothesis d=0 are 0.511, 0.386 and 0.322. In conclusion the CRDW test indicates the variables are co-integrated as the d-statistic of 0.5947 is above the critical values at all significance levels.

5.1.2 Augmented Engle-Granger (AEG) Test

An alternative test of co-integration among variables is the Augmented Engle-Granger (AEG) Test. The maximum likelihood techniques are applied to test the estimated residual of the co-integration regression shown in the CRDW test and is represented by the Augmented Dickey-Fuller tau statistic with zero mean:

$$\log GDP_t = \beta_1 + \beta_2 \log UNEMP_t + \beta_3 \log CPI_t + \beta_4 \log INT_t + \mu_t$$

Each variable in itself is non-stationary as proved in the stationarity tests previously; however through a unit root test the residuals of the co-integrating regression might be stationary; hence the above regression model is performed and the residuals are extracted. From these the delta residuals i.e. the one-period-change in the value of the residuals are regressed on the one-period-lagged residuals. From that the Augmented Engle-Granger statistic d is found.

---

70 Gujarati, 2003, page 823-824
71 Koop 2008, page 224-225
72 See appendix 4 page 48
\[ \Delta \hat{\mu}_t = -0.2968 \hat{\mu}_{t-1} \]
\[ SE = (0.0725) \]
\[ t = (-4.09) \]
\[ R^2 = 0.1471 \]
\[ d = 1.9850 \]

The null-hypothesis of the test is:

\[ H_0: \text{No co-integration} \]
\[ H_1: \text{Co-integration} \]

The Engle-Granger tests 5% critical value is -3.39, given no intercept nor deterministic trend\(^73\) and with \(T=100\) being the appropriate measure closest to the period of our time series from 1984:1-2008:3. Hence, the AEG \(t\) value of -4.09 is more negative than the critical value showing that co-integration do exist among the four variables.

### 5.1.3 Pair wise co-integration tests of variables

Equivalent co-integration tests have been performed pair-wise among the variables. In spite of all expectations none of the tests indicate a co-integrating relationship between the variables one-on-one, which contradict the original expectations and is a source of some consideration.

In order to clarify the issue of pair-wise co-integration, take a closer look at the regression function between \(GDP_t\) and \(\text{Interest Rate}_t\)\(^74\):

\[
\log GDP_t = \beta_1 + \beta_2 \log INT_t + \mu_t
\]
\[
\log GDP_t = 7.7564 - 1.0883 \log INT_t
\]

\(^73\) Given that error terms are assumed to have a zero mean the intercept term is often not included. Koop, 2008 page 221

\(^74\) See appendix 4 page 49
The Durbin-Watson d statistic of 0.3852 is bordering to the CRDW critical value of 0.386, which statistically incur that co-integration between GDP_t and INT_t statistically can be rejected. Equivalently the outcome of the Engle-Granger co-integration test is as follows:

\[ \Delta \mu_t = -0.1906 \mu_{t-1} \]
\[ t = (-3.17) \]

The estimated t statistic from the residuals based on the GDP_t and INT_t regression is within the borderline of the Engle-Granger 5% critical value of -3.39, thus the non-co-integration test cannot be rejected.

This is the overall picture of the pair-wise co-integration tests. The anticipated signs of pair-wise co-integration relationship did not occur; instead several of the variables are at the borderline of being statistically significant. An economic interpretation of these tests is as mention previously that even though a long run equilibrium relationship is present in a common model consisting of all the explanatory variables, no single pair-wise regression can explain this relationship. The direct relationship is thus dependent on third-factor influences for co-integration to occur.

The long run equilibrium relationship between all four explanatory variables must be corrected for in the overall multivariate model before continuing analyzing on the impact on the US real estate transaction-based returns. Consequently an error correction model is to be developed.

---

75 See appendix 4 page 50
5.2 The error correction model 1

The properties of the error correction model mean the risk of a spurious regression is no longer present. As shown all variables have unit roots whereas they are stationary when integrated once I(1). Given the co-integration properties of the variables the equilibrium error is also stationary. Including an extra parameter of the one-period-lagged error term from the co-integration regression imply all variables are stationary and OLS estimation is still applicable for future model testing.

The error term to be added as an explanatory variable in the multivariate model through the co-integration regression is as follows:

\[ \varepsilon_{t-1} = \log GDP_{t-1} - \beta_1 - \beta_2 \log UNEMP_{t-1} - \beta_3 \log CPI_{t-1} - \beta_4 \log INT_{t-1} \]

The estimation of the error term is available from paragraph 5.1.1. Finally we can estimate the multivariate model to be used for the actual econometric analysis based on the error correction model (ECM) above. Including the extra error term in the model in addition to the lagged values of the explanatory variables due to the I(1) property, the distributed lag model 1 to be used going forward in this paper is:

\[ \Delta TBI_t = \beta_1 + \beta_2 \Delta UNEMP_t + \beta_3 \Delta UNEMP_{t-1} + \beta_4 \Delta CPI_t + \beta_5 \Delta CPI_{t-1} + \beta_6 \Delta INT_t + \beta_7 \Delta INT_{t-1} + \beta_8 \Delta GDP_t + \beta_9 \Delta GDP_{t-1} + \lambda \varepsilon_{t-1} + \mu_t \]

Where \( \Delta X_t = \log X_t - \log X_{t-1} \)

The model is a function of delta TBI\(_t\), with intercept \( \beta_1 \), the explanatory variables in its original form and lagged one period due to the property of I(1), the long run equilibrium error \( \varepsilon_{t-1} \) to correct for co-integration in the model and finally an error term \( \mu_t \).

---

76 Koop 2008, page 224
77 Koop 2008, page 225
78 Koop, 2008, page 227
Economically the intuition behind the model is that the US real estate returns are expected to be influenced from the change in the four macroeconomic variables (i) unemployment, (ii) inflation, (iii) the long term interest rate and (iv) naturally the growth in the US economy. In addition to that the returns are also under influence of the changes in the macro economy the previous quarter i.e. the one-period lagged variables. An estimation of the model will clarify, to what extend the returns have a one-quarter delayed reaction on the changes on the general economy. This also makes good economic sense because real estate as mentioned in the beginning of the thesis is an asset with only quarterly valuations, thus it may have some delays to it – also, as the transaction-based return used in this paper is unsmoothed values i.e. being corrected for moving average properties, no autoregressive lag of the dependent variable is needed in the error correction model as it ought not to suffer from any time lagged influences. The t-values of the error correction model will clarify if the time-lagged explanatory variables are significant or actually equal to zero.

Finally, the inter-dependence of the macro economy highlighted in the co-integration regression, which increase the risk of spurious estimation of the model is adjusted for by adding an extra parameter of the long-run equilibrium error obtained from the co-integration regression to model 1.
Chapter 6
Classical Linear Regression Model Assumptions

By now the specification of the multivariate dynamic error correction model has been defined on the basis of the stationarity and co-integration tests. The model will be the actual starting point for the econometric analysis of the macro economic influence on the US real estate return index. So far OLS estimates of the models have been derived without considering the assumptions of the classical linear regression model (CLRM). However to obtain the desirable properties i.e. for the Gauss-Markov theorem of best linear unbiased estimators (BLUE) to apply to the model, the OLS estimates require fulfilling of the CLRM assumptions. Empirical regression analyses do not always fulfill the classical assumptions, thus it is important to check the assumptions through hypothesis testing. If the model assumptions is violated the economic insight of the model may at best be inefficient or even worse be seriously biased or misleading. Hence, applicable estimators must be developed in case any assumptions occur to be false.

The CLRM given by Gujarati (2003) assumptions that are to be tested are as follows:

1. The regression model is linear in the parameters
2. The values of the regressors, the X’s are fixed in repeated sampling
3. For given X’s, the mean value of the disturbances μ_t is zero
4. For given X’s, the variance of μ_t is constant or homoscedastic
5. For given X’s, there is no autocorrelation in the disturbances

---

79 Koop, 2008, page 121-122
80 Gujarati, 2003, page 335: The assumptions of the classical linear regression model
6. If the X’s are stochastic, the disturbance term and the X’s are independent or at least uncorrelated
7. The number of observations must be greater than the number of regressors
8. There must be sufficient variability in the values taken by the regressors
9. The regression model is correctly specified
10. There is no exact linear relationship (i.e. multicollinearity) in the regressors
11. The stochastic disturbance term \( \mu_t \) is normally distributed

The following pages will test above assumptions, with primary focus on the normality assumption, autocorrelation and homoscedasticity of the time series.
Chapter 7
Model Specification

Before analyzing the model developed in the previous sections it is crucial to assure normal distributed error terms. Three actions will be taken to assure the normality assumption: (i) a histogram of the residuals, (ii) normal probability plot and (iii) execution of the Jarque-Bera numerical test. As previously stated the multivariate error correction model 1 is, all values being in logs:

\[ \Delta TBI_t = \beta_1 + \beta_2 \Delta UNEMP_t + \beta_3 \Delta UNEMP_{t-1} + \beta_4 \Delta CPI_t + \beta_5 \Delta CPI_{t-1} + \]
\[ \beta_6 \Delta INT_t + \beta_7 \Delta INT_{t-1} + \beta_8 \Delta GDP_t + \beta_9 \Delta GDP_{t-1} + \lambda \varepsilon_{t-1} + \mu_t \]

Where \( \Delta X_t = \log X_t - \log X_{t-1} \)

The estimates of the model are81:

\[ \Delta TBI_t = 0.000363 - 0.0943 \Delta UNEMP_t + 0.0104 \Delta UNEMP_{t-1} - 0.0183 \Delta CPI_t - \\
0.0022 \Delta CPI_{t-1} - 0.037 \Delta INT_t - 0.002034 \Delta INT_{t-1} + 0.0508 \Delta GDP_t + \\
1.4309 \Delta GDP_{t-1} + 0.0516 \varepsilon_{t-1} \]

\begin{align*}
\text{SE} & = (0.00705) \quad (0.0547) \quad (0.0545) \quad (0.0199) \quad (0.9256) \\
& = (0.0199) \quad (0.0492) \quad (0.0478) \quad (0.8524) \quad (0.9256) \\
\text{t} & = (0.05) \quad (-1.72) \quad (0.19) \quad (-0.92)
\end{align*}

81 See appendix 5 page 63
The estimation of the model does not appear promising; the variables explain 12.6% of the variation in $\Delta TBI$, however none of the variables are significant according to the appropriate t-statistics. Most noticeably the co-integration error coefficient ought to be significant in the presence of co-integration in the model. Performing an F-test will assess the overall significance of the model. The result is an F-value of 1.3776$^{82}$, which cannot reject that the model equals zero given its appropriate critical value i.e. it does not explain anything of $\Delta TBI$. The normality test will clarify the issues of the model and what may be done to solve the problems. Next is a graphical inspection of the residuals of the model$^{83}$. It is clear from the figures that there are some indications on non-normality in the behavior of the residuals. The histogram has some skewness to it, while the probability plot ought to follow the diagonal line to be normal, but instead confirms the findings from the histogram, i.e. that the residuals are not symmetrically distributed, but have too many large errors below the diagonal line. Meanwhile, the plot also indicates the residuals have excessive kurtosis to it, especially the upper left outlier in the plot. A numerical test as that of the Jarque-Bera test should be able to confirm the findings.

$^{82}$ La Cour, 2006, Lecture Notes 1 slide 29: $F = \frac{R^2/(k-1)}{[1-R^2]/(n-k)} - F(k-1,n-k)$ See appendix 9 page 121

$^{83}$ See appendix 5 page 65
7.1 Jarque-Bera normality test

The Jarque-Bera test is an asymptotic test based on the OLS residuals. The value is found from the regression function in SAS based on the following formula, where $S$ is the skewness coefficient and $K$ the Kurtosis coefficient\(^{84}\).

\[
JB = n \left[ \frac{S^2}{6} + \frac{K - 3^2}{24} \right]
\]

The JB statistic follows the chi-square distribution with 2df, thus the value of 37.1850 i.e. a p-value < 0.0001 rejects the normality assumption. Possible solutions to the model are either to redo the model specification according to the classical linear regression model assumption \(^{95}\) or look at potential outliers. However even if redoing the functional form of the model\(^{86}\), leaving out one or more explanatory variables from the model will risk a specification error of omitting a relevant variable\(^{87}\) and the co-integration paragraph supports inclusion of the variables in its original and lagged form. As for the functional form log-values is typically how macroeconomic data is used in econometric analysis.

7.2 Ramsey’s Regression Specification Error Test

A general way of testing for misspecification of the model is the Ramsey’s regression specification error test (RESET test). Even though it is not helpful in choosing a better alternative, it can be applied to test the specification of the selected model. The null-hypothesis and the F-test is as follows\(^{88}\):

\[ H_0: \text{no misspecification} \]

---

\(^{84}\) Gujarati, 2003, page 148

\(^{85}\) Gujarati, 2003, page 506

\(^{86}\) Gujarati, 2003, page 175 and 508-509

\(^{87}\) Gujarati, 2003 page 508-509

\(^{88}\) Gujarati, 2003, Ramsey’s RESET test of Specification Error, page 521-523
H₁: not H₀

\[ F = \frac{(R_{new}^2 - R_{old}^2) / \text{number of new regressors}}{(1 - R_{new}^2) / n - \text{number of parameters in the new model}} \]

The test statistic can be extracted from the SAS output, and as seen from below table the RESET test is insignificant at all power levels indicating there should not be any problems with the specification of the model, in spite of the insignificance problem of all explanatory coefficients.

<table>
<thead>
<tr>
<th>Power</th>
<th>RESET</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.0054</td>
<td>0.9418</td>
</tr>
<tr>
<td>3</td>
<td>0.5862</td>
<td>0.5587</td>
</tr>
<tr>
<td>4</td>
<td>0.4055</td>
<td>0.7495</td>
</tr>
</tbody>
</table>

Table 7.1 Ramsey’s Specification Test of Model 1

Further, also the model’s Durbin-Watson d-statistic of 1.9604 does not indicate any correlation in the residuals, which also indicate an acceptable specified model. Hence, the non-normality of the model must be explained from another side.

---

89 See appendix 5 page 63
7.3 Outliers among residuals

Looking towards the residuals of the time series\textsuperscript{91} it is apparent that certain outliers are present, and they may be the cause of the non-normal distribution.

Draper and Smith warn about rejecting outliers\textsuperscript{92}, as the outliers may provide useful information about the values of some of the coefficients. Meanwhile if the outliers can be explained through unique events that are not likely to be repeated, it may be sensible to remove them from the model. From the residual plot I have identified the most distinct outliers which are the following year and quarter, respectively:

- 1987:4
- 1988:4
- 1991:4
- 1992:4
- 2005:2
- 2007:4

Notice the outliers are mainly fourth quarter values. This may partly be explained by some seasonal effect on the real estate data; however there are certain historic macroeconomic events that lie behind the outliers in question. The events will be described shortly and I would argue the uniqueness of these events makes it reasonable to remove some or all of

\textsuperscript{91} See appendix 5 page 64
\textsuperscript{92} Gujarati, 2003, page 541
the outliers in question. Note that TBI is only measured quarterly; hence it can be regarded as a sort of delayed response function, where events influence the time series one period lagged.

### 7.3.1 Reasoning behind removal of outliers

1) **Black Monday, October 19, 1987**[^3]  

In 1987 stock markets drop hit the United States resulting in huge drops of values in very short time periods. The Dow Jones dropped 22.6% in a single day followed by an overall US stock market knock down of 23% in just two days. In the tale of the global stock market panic, it is apparent from my dataset the interest rate rose significantly and in spite of a soft landing of the US economy in 1986 with an inflation drop, the CPI suddenly suffered from a rapid increase. The event which caused the sudden deviations in the macro economic variables is causing an immediate influence on the regression model resulting in the major outlier in fourth quarter 1987 and 1988.

2) **The after-comings of the 1990-1991 recessions**[^4]  

In the tale of the stock market crisis in 1987, the economy and stock market began to recover. However, a potential collapse of savings and loans caused a condition of panic and lead to a sharp recession in the United States and other major economies as United Kingdom and Canada. In 1992 the recession had ended and the economic stagnation lessened the level of inflation from a high level and also the unemployment rate had peaked and employment rate grew progressively.

3) **The 2005 United States housing bubble**[^5]  
[Outlier 2005:2]

[^3]: Blanchard 2.ed, 2000, page 301 and  
In 2005 the economic event in the US had a direct influence on the real estate market because of the so-called housing bubble. Housing bubbles are typically a result of rapid increases in valuations of the underlying properties until they reach an unsustainable level and a market correction occurs. Hereafter, the property owners finds themselves in a position where their mortgage debt is higher than the actual value of their property. The effect of the housing bubble is directly mirrored in the US transaction-based index of real estate returns after a few quarters of abnormal positive deviation in the return index, dropping to its natural level again.


The real estate market peaked in 2005 and began a correction of the market. A forecast of a free fall of the housing market and a potential recession to come had already been warned about, and in March 2007 the US sub-prime mortgage industry collapsed due to complex and risky debt markets, especially the collateralized debt obligation market[^98].

The victims of the crisis involved big companies as lenders Northern Rock and Wachovia, American insurance company AIG, banks as Bear Stern, Lehman Brothers, and Goldman Sachs etc. and not at least the bankruptcies of Government sponsored enterprises Fannie Mae and Freddie Mac. This major event can obviously directly be seen from TBI, experiencing significant negative growth for the first time since the primo 1990 recession.

In addition to above mentioned event, one could question mark, how come the dot-com bubble in the late 1990s is not an issue on equal terms as the historical events mentioned. Obviously the IT bubble does not have a direct influence on the real estate market, but secondly and most important the bubble covered roughly the years 1995-2001 i.e. it was a long term crises and not a single event visible from our data. Hence, no outliers were detected in the time period and it does not cause any trouble for the normality of our model.

5) Conclusion on outlier removal

[^97]: Danske Analyse- [www.danskeanalyse.danskebank.dk](http://www.danskeanalyse.danskebank.dk)
In conclusion it is reasonable to believe the above-mentioned events are providing useful information to our coefficients, and there is no doubt of the importance of the events. Nevertheless I would claim the uniqueness of the events allows me to remove the outliers that violate the normality assumption in order to make a proper long term analysis of the US real estate return index. One should however bear in mind the heavy volatility in the market in those crucial moments.

7.3.2 Removal of outliers and model selection

From the residual plot I have, as discussed, identified the most distinct outliers of the model, and consequently adjusted the model. The adjustment resulted in three different adjusted error correction models in addition the original model 1 – all being normal distributed according to the Jarque-Bera test. The adjusted models contain the removal of five of the six outliers defined and the selection procedure between the models will be done according to the Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC)\(^9^9\). Hence, the theory says the model with the lowest value of AIC and SIC is preferable. The values of the models are shown below:

<table>
<thead>
<tr>
<th>Model selection</th>
<th>Original ECM 1</th>
<th>Adj ECM 2</th>
<th>Adj ECM 3</th>
<th>Adj ECM 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outliers removed (Year; Quarter)</td>
<td>-</td>
<td>1987:4</td>
<td>1987:4</td>
<td>1987:4</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>1988:4</td>
<td>1988:4</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1991:4</td>
<td>1991:4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2007:4</td>
<td>-</td>
<td>-</td>
<td>2007:4</td>
</tr>
<tr>
<td>AIC</td>
<td>-514.21094</td>
<td>-539.9548</td>
<td>-539.74169</td>
<td>-538.83524</td>
</tr>
<tr>
<td>SIC</td>
<td>-488.56746</td>
<td>-514.84624</td>
<td>-514.63309</td>
<td>-513.72665</td>
</tr>
</tbody>
</table>

Table 7.2 Error Correction Model Selection via AIC and SIC

\(^9^9\) Gujarati 2003, page 537
The removal of outliers are based on the residuals plot and the economic events simply by removing the detected outliers from the excel data sheet before performing the regression analysis. A more econometric approach would be to base the model selection on a dummy regression analysis. From appendix 5 can be seen such dummy regression models\textsuperscript{100}, including a dummy variable 1 at time $t$ of the outlier and a dummy variable -1 at time $t+1$ of the outlier, in order to assure immediate adjustment of the delta values of the model. The dummy regressions confirm the findings of the model selection above. Given the validation of the model selection the analysis are based on the simple removal approach.

Table 7.2 clearly shows an improvement of the model by removing some of the most distinct outliers as AIC and SIC become more negative regardless of which adjusted model is chosen. Further, the model comparison reveals the importance of assessing all the events when removing outliers, as a result the adjusted error correction model 2 seems to be the preferred model having the most negative AIC and SIC values. In addition it is worth noticing the coefficient of the co-integration error term is only significant in the selected model 2\textsuperscript{101}. This corresponds to my prior mentioned expectations, as significance of the co-integration error coefficient ought to be present, if co-integration applies to the model.

It is nevertheless plausible to claim that removing five outliers is a high number and may jeopardize the strength and power of the model. However in spite of a potentially weakened model, the removals are necessary in order to obtain a statistically useful normal distributed multivariate regression model. Consequently going forward the new adjusted error correction model will be the starting point of the analysis to come.

\textsuperscript{100} See appendix 5 page 76-79
\textsuperscript{101} See appendix 5 page 66-75 for estimation of the three adjusted error correction models
For the sake of visualization the actual changes to the delta TBI time series can be seen in the graph, when the distinct outliers have been removed: The new error correction model – model 2 - has a much better looking pattern over time, with less volatility and an immediate interpretation of the model would also be an anticipation of no autocorrelation in the residuals. This will be analyzed further in coming paragraphs.

### 7.4 The adjusted error correction model 2

The new multivariate error correction model adjusted for outliers caused by unique economical events is\(^ {102}\): All values are in logs.

\[
\Delta TBI_t = \beta_1 + \beta_2 \Delta UNEMP_t + \beta_3 \Delta UNEMP_{t-1} + \beta_4 \Delta CPI_t + \beta_5 \Delta CPI_{t-1} + \\
\beta_6 \Delta INT_t + \beta_7 \Delta INT_{t-1} + \beta_8 \Delta GDP_t + \beta_9 \Delta GDP_{t-1} + \lambda_2 \varepsilon_{t-1} + \mu_t
\]

Where \(\Delta X_t = \log X_t - \log X_{t-1}\)

The estimates of model 2 are:

\[
\hat{\Delta TBI}_t = 0.000664 - 0.1040\hat{\Delta UNEMP}_t - 0.007546\hat{\Delta UNEMP}_{t-1} - 0.008858\hat{\Delta CPI}_t - \\
0.004875\hat{\Delta CPI}_{t-1} - 0.0688\hat{\Delta INT}_t + 0.00954\hat{\Delta INT}_{t-1} + 0.2448\hat{\Delta GDP}_t + \\
1.3453\hat{\Delta GDP}_{t-1} + 0.0519 \varepsilon_{t-1}
\]

\(^{102}\) See appendix 5 page 66
The estimated coefficients of model 2 now show a different picture than the original model. The explanatory power of the variation in $\Delta TB_I$ is as much as 22.27%. However, the relatively high $R^2$ and only few significant variables may indicate multicollinearity amongst the variables. It seems, however, from the model that changes in employment rate does significantly affect the return on real estate investments in the US, and the same is the matter for the time lagged economic growth GDP. More important change to the model is the now significant co-integration error term. The positive error correction term implies the change in delta $TB_I$ is below its equilibrium level, and the error term will cause the change in the dependent variable to be positive, meaning the level of the transaction-based return will rise in period $t$.

Once again an F-test is performed with the number of explanatory variables $k = 10$ and the number of observations $n = 91$, to assess the overall significance of the model. The F-value calculated is 2.5785$^{103}$, which is above its appropriate 5% critical value of 1.96 ($n = 60$) or 2.04 ($n = 120$)$^{104}$, as a result the model is statistically significant.

---

$^{103}$ See appendix 9 page 121
$^{104}$ Gujarati, 2003, page 966 – F distribution
The histogram has also improved even though it still has some skewness to it, whereas the probability plot of model 1’s residuals follows the diagonal line with fewer large errors implying a normal distributed model\(^ {105}\). The numerical Jarque-Bera normality test referred to earlier confirms the findings with a test statistic of 5.2210 and p-value of 0.0735, hence we cannot reject the normality assumption. Lastly, Ramsey’s RESET test also indicates a well specified model with insignificant values across all levels.

Looking towards the residual plot from model 2 regression\(^ {106}\), the residuals are better distributed on both sides of zero, with lower volatility and less significant outliers that seemingly does not cause any problems according to the normality tests above.

\(^{105}\) See appendix 5 page 69
\(^{106}\) See appendix 5 page 68
Chapter 8
Time Series Volatility

A special characteristic of financial time series also present here, is the fact that the time series variables are random walks in their level form, but I(1) stationary. However, taking the first difference and working with changes in log values often imply conditional variance of the variance i.e. error term volatility. In these cases Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models come in handy.\footnote{Gujarati, 2003, page 856}

ARCH models were introduced by Engle (1982) and GARCH by Bollerslev (1986) and are estimated by the Maximum likelihood method under the assumption that the errors are conditionally normally distributed\footnote{La Cour, 2006, ARCH Models lecture notes}. In case of heteroscedastic behavior of the OLS coefficients, the model will violate our CLRM assumptions and will no longer have minimum variance i.e. the obtained estimates will no longer be BLUE (Best Linear Unbiased Estimator)\footnote{Gujarati, 2003, page 393-394}, however the estimates will still be linear and unbiased.

The purpose of ARCH models is to model (and forecast) conditional variances and to manage potential heteroscedasticity in error terms in order to obtain more efficient estimates of the model. The variance of a dependent variable is modeled as a function of past values of the dependent variable and exogenous variables. Being so closely related to the AR (p) process, it is not my anticipation to find any ARCH effect in the error correction
model for the dependent variable, as model 2 does not include lagged values of the dependent values, but only of the explanatory variables. Also, the fact that the time series data are applied only on a quarterly basis implies that any heteroscedastic behavior is less likely to appear in the regression model.

The ARCH model with p lags measures the volatility from the previous periods via the lag of the squared residual from the mean equation i.e. the conditional variance is given by past squared errors is\(^{110}\):

\[
\sigma^2 = \gamma_0 + \gamma_1 \mu_{t-1}^2 + \gamma_2 \mu_{t-2}^2 + \ldots + \gamma_p \mu_{t-p}^2
\]

Where the gammas are the coefficients estimate by e.g. maximum likelihood. As mentioned the model tests for autocorrelation in the error variance terms and the number of lags depends on the significance of the coefficient. In case additional lags do not add significant explanatory power to the model they can be ignored. An extension to the model is the generalized ARCH model i.e. GARCH. The model assumes the conditional variance is dependent on both the volatility of the previous period (the ARCH term) and the forecasted variance from the last period (the GARCH term). The model is as follows\(^{111}\):

\[
\sigma^2 = \gamma_0 + \gamma_1 \mu_{t-1}^2 + \ldots + \gamma_p \mu_{t-p}^2 + \lambda_1 \sigma_{t-1}^2 + \ldots + \lambda_p \sigma_{t-p}^2
\]

The properties of the GARCH model are similar to those of the ARCH model; however it is more flexible and is typically applied as a replacement for of higher-order ARCH models. In addition hereto there exist a number of other equivalent models that will not be described here\(^{112}\).

\(^{111}\) Koop, 2008, page 204-205
\(^{112}\) Koop, 2008, page 204
8.1 ARCH test on model 2

Going forward a number of tests for ARCH effects will be performed on our stationary multivariate model 2. First a graphical assessment of our dependent variable will indicate the potential of heteroscedasticity in the times series. In addition hereto Engle’s ARCH tests will be performed:

- ACF / PACF graphical outlook
- Engle’s ARCH test based on an auxiliary regression
- LM test for ARCH disturbances (Engle’s test in SAS)

Financial time series are often characterized of having ARCH effects, however given the properties of model 2 I do not necessarily anticipate ARCH effects in the model. First of all, even though the macroeconomic variables independently might have time varying variance in their original values, the multivariate model influencing the heavy asset class of real estate is not expected to be affected. Second, the functional log form of model 2 as well as the lack of any autoregressive processes of our dependent variable in the model opposes the risk of heteroscedastic behavior in the time series residuals. Lastly, the removal of the outliers to obtain normality of the model has also somewhat lessened the volatility of the model.

113 Gujarati, 2003, page 859 and La Cour, 2006, Exercise 8 – ARCH and GARCH
8.1.1 Graphical indication of ARCH effects

Indications of ARCH effects in the model can be seen from obtaining the squared residuals of our regression model 2\(^{114}\). In theory the lag of \( p \) is the last significant spike from the PACF function of squared residuals. The PACF only have minor indications of an ARCH (\( p \)) process, as borderline values occur at lag 6 and 16\(^{115}\). It is however difficult to correctly interpret the graphical overview, hence the numerical tests are more compelling.

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
<th>Autocorrelations</th>
<th>Partial Autocorrelations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4.076E-8</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-4.883E-9</td>
<td>-1.1180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-6.008E-13</td>
<td>-0.00001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-7.598E-10</td>
<td>-0.01864</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.34136E-9</td>
<td>0.08196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-8.577E-10</td>
<td>-0.02104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7.075E-9</td>
<td>0.17356</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-8.272E-12</td>
<td>-0.00020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>-2.287E-10</td>
<td>-0.00561</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8.1 ACP of PACF of Squared Residuals from Model 2. See all lags from appendices

PACF function of squared residuals. The PACF only have minor indications of an ARCH (\( p \)) process, as borderline values occur at lag 6 and 16\(^{115}\). It is however difficult to correctly interpret the graphical overview, hence the numerical tests are more compelling.

8.1.2 Engle’s ARCH test

The ARCH test tests the joint significance of the estimated squared residuals at different lags, and is based upon the residuals obtained from model 2.

\[
\Delta TBI_t = \beta_1 + \beta_2 \Delta UNEMP_t + \beta_3 \Delta UNEMP_{t-1} + \beta_4 \Delta CPI_t + \beta_5 \Delta CPI_{t-1} + \beta_6 \Delta INT_t + \beta_7 \Delta INT_{t-1} + \beta_8 \Delta GDP_t + \beta_9 \Delta GDP_{t-1} + \lambda \varepsilon_{t-1} + \mu_t
\]

\(^{114}\) La Cour, 2006, Exercise 8 – ARCH and GARCH page 4

\(^{115}\) See appendix 6 page 81 for view of PACF lag 16.
From that an auxiliary regression of the estimated squared residuals is run:

\[ \hat{\mu}_t^2 = \gamma_0 + \gamma_1 \hat{\mu}_{t-1}^2 + \ldots + \gamma_p \hat{\mu}_{t-p}^2 + error_t \]

The test statistic of the auxiliary regression is \( Q = n*R^2 \sim \chi^2_{0.95(p)} \), and the null-hypothesis is:

\[ H_0 = \gamma_0 = \gamma_1 = \ldots = \gamma_p = 0 \quad \text{(No ARCH)} \]
\[ H_1 = \text{not } H_0 \]

The lag length of the ARCH models can be selected either through an information criterion or simply until a coefficient equals zero. Therefore, the initial auxiliary regression performed has a lag length of \( p = 1 \). The OLS estimates from the maximum likelihood regression are\(^{116}\):

\[ \hat{\mu}_t^2 = 0.000131 - 0.1229 \hat{\mu}_{t-1}^2 \]

<table>
<thead>
<tr>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5.65)</td>
<td>(&lt;.0001)</td>
</tr>
<tr>
<td>(-1.24)</td>
<td>(0.2179)</td>
</tr>
</tbody>
</table>

The test statistic calculated from the auxiliary regression \( Q = 89*0.0172 = 1.5308 \) is undeniably below its 95% chi-square critical value with 1 degree of freedom of 3.84146 indicating no ARCH effect to the model. Also the insignificant p-value indicates the estimated lagged squared residual does not add any significant explanatory power to the model and can be ignored. The null-hypothesis of no ARCH can neither be rejected if testing for \( p = 3 \) lags\(^{117}\). However, the PACF function did have some indication of higher-order ARCH effects at \( p = 6 \), hence an equivalent test could add some value to the understanding of the model, and this is demonstrated next from the automatically derived statistic in SAS.

---

\(^{116}\) See appendix 6 page 83
\(^{117}\) See appendix 6 page 84 for \( p = 3 \) regression output
8.1.3 LM test for ARCH disturbances

From the model 2 regression function equivalent test statistics for ARCH disturbances can be obtained in SAS for higher-order lags\textsuperscript{118}, and the values clearly shows no signs of time varying variance as the no ARCH effect hypothesis can be accepted also of higher-order p, given the value of the LM test statistics. The rejection of any ARCH effects in the model also assures no GARCH is present, thus the effect will not be further assessed. In conclusion the specific characteristics of model 2 caused the ARCH tests to oppose usual behavior of financial time series; hence there is no need to make any additional adjustments to the model. Further, given the homoscedastic properties of model 2, the OLS coefficients have a minimum variance i.e. of the model is still BLUE.

<table>
<thead>
<tr>
<th>Order</th>
<th>Q</th>
<th>Pr &gt; Q</th>
<th>LM</th>
<th>Pr &gt; LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.3495</td>
<td>0.2454</td>
<td>1.9147</td>
<td>0.1664</td>
</tr>
<tr>
<td>2</td>
<td>1.3495</td>
<td>0.5093</td>
<td>2.0739</td>
<td>0.3545</td>
</tr>
<tr>
<td>3</td>
<td>1.3829</td>
<td>0.7095</td>
<td>2.3338</td>
<td>0.5061</td>
</tr>
<tr>
<td>4</td>
<td>2.0364</td>
<td>0.7291</td>
<td>2.5828</td>
<td>0.6299</td>
</tr>
<tr>
<td>5</td>
<td>2.0799</td>
<td>0.8380</td>
<td>2.6305</td>
<td>0.7567</td>
</tr>
<tr>
<td>6</td>
<td>5.0791</td>
<td>0.5337</td>
<td>5.0090</td>
<td>0.5427</td>
</tr>
<tr>
<td>7</td>
<td>5.0979</td>
<td>0.6480</td>
<td>5.0094</td>
<td>0.6588</td>
</tr>
<tr>
<td>8</td>
<td>5.3154</td>
<td>0.7234</td>
<td>5.5089</td>
<td>0.7021</td>
</tr>
<tr>
<td>9</td>
<td>5.3215</td>
<td>0.8054</td>
<td>5.6197</td>
<td>0.7773</td>
</tr>
<tr>
<td>10</td>
<td>6.1780</td>
<td>0.8001</td>
<td>5.7559</td>
<td>0.8353</td>
</tr>
<tr>
<td>11</td>
<td>6.4746</td>
<td>0.8399</td>
<td>6.0303</td>
<td>0.8713</td>
</tr>
<tr>
<td>12</td>
<td>6.5390</td>
<td>0.8865</td>
<td>6.1187</td>
<td>0.9100</td>
</tr>
</tbody>
</table>

Table 8.2 Model 2 ARCH test for higher-order p

\textsuperscript{118} See appendix 6 page 85
Chapter 9
Autocorrelation

Given no ARCH in the multivariate regression function, it is now possible to test for autocorrelation in the time series model. Autocorrelation tests has previously been performed in paragraph of stationarity of the time series, however in order to fulfill assumption 5 in of the linear regression model the tests will shortly be mentioned also for model 2. It is my expectations that prior changes to the model, i.e. the inclusion of lagged values of the explanatory variables in the model to obtain stationarity as well as including the co-integration term and the removal of distinct outliers, ought to result in no problems of autocorrelation for model 2.

Firstly, the Durbin-Watson test examines the presence of autocorrelation in the model. In doing the numerical autocorrelation test, the d statistic underlies certain assumptions. As opposed to previous Durbin-Watson tests performed in this paper it is now certain that the error term is normally distributed subject to model 2. Further the co-integration tests resulted in a model of lagged explanatory variables; whereas the lagged value of the dependent variable is not included in the model, hence model 2 fulfill the assumptions of the Durbin-Watson test.

\(^{119}\) In addition to the mentioned assumptions, the model is assumed to fulfill (1) model includes an intercept term, (2) explanatory variables are non-stochastic, and (3) disturbance terms are generated by first-order autoregressive scheme, and cannot be used to detect higher-order autoregressive schemes. Gujarati, 2003, page 467
\(^{120}\) Gujarati, 2003, page 467-468 - Durbin-Watson assumptions
The $d$ statistic is given from the estimation output of model 2 in testing for normality\textsuperscript{121}. The $d$ statistic of 1.8711\textsuperscript{122} lies within the significance points of the Durbin-Watson test\textsuperscript{123}. Hence, there are no indications of autocorrelation in the model. Also, The Breusch and Godfrey's autocorrelations test below shows no signs of serial correlation of any order, as all test statistics in the Lagrange Multiplier test are insignificant on a 5% significance level.

<table>
<thead>
<tr>
<th>Godfrey's Serial Correlation Test</th>
<th>LM</th>
<th>Pr &gt; LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.0105</td>
<td>0.9183</td>
</tr>
<tr>
<td>AR(2)</td>
<td>1.4654</td>
<td>0.4806</td>
</tr>
<tr>
<td>AR(3)</td>
<td>2.5206</td>
<td>0.4716</td>
</tr>
<tr>
<td>AR(4)</td>
<td>5.1018</td>
<td>0.2770</td>
</tr>
</tbody>
</table>

Table 9.1 Model 2 Breush-Godfrey Serial Correlation test\textsuperscript{124}

Lastly, a graphical investigation through a line plot of the residuals of the model against time, and equivalently a scatter plot of the standardized residuals against the lagged values of the standardized residuals confirm the expectations to the model\textsuperscript{125}. The residuals of the graph are placed randomly around zero, thus the observations do not appear to be correlated. The residuals are random i.e. statistically independent.

\textsuperscript{121} See model 2 normality test paragraph 7.4

\textsuperscript{122} Gujarati, 2003, page 467 - Durbin-Watson $d$ statistic is given by: $d = \frac{\sum_{t=2}^{n}(\hat{\mu}_t - \hat{\mu}_{t-1})^2}{\sum_{t=1}^{n-1} \hat{\mu}_t^2}$.

\textsuperscript{123} Gujarati, 2003, page 970 – Durbin-Watson $d$ statistic

\textsuperscript{124} See appendix 7 page 87

\textsuperscript{125} See appendix 7 page 88
Chapter 10
Model Estimation

By now we have cleared the model for any non-normal behaviour and tested for assumption violations of homoscedasticity and autocorrelation in the multivariate time series model. Still, the estimation of model 2 evidently had many insignificant coefficients, in spite of a reasonable $R^2$ and significant explanatory power, which can be a sign of multicollinearity. Given the co-integration properties the model ought to stay clear of any collinearity problems through the error correction term. Nevertheless, if the explanatory variables are scrutinized we see a high correlation between particularly interest rate and the GDP\(^{126}\), which is confirmed through a scatter plot of the two variables against each other showing an almost exact linear relationship\(^{127}\).

![Figure 7.8: Log GDP vs Log INT](image)

<table>
<thead>
<tr>
<th>Correlations (log values)</th>
<th>TBI</th>
<th>UNEMP</th>
<th>CPI</th>
<th>INT</th>
<th>GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBI 1</td>
<td>-0.21841</td>
<td>-0.26312</td>
<td>-0.8525</td>
<td>0.954163</td>
<td></td>
</tr>
<tr>
<td>UNEMP 1</td>
<td>0.013174</td>
<td>0.044752</td>
<td>-0.19564</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPI 1</td>
<td>0.427518</td>
<td>-0.28119</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT 1</td>
<td>-0.91084</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 10.1 Correlations of the log values of the variables

\(^{126}\) See appendix 8 page 94-95

\(^{127}\) For covariance and correlation among the explanatory variables, see appendix 1 page 10
These properties inevitably do seem like multicollinearity\(^{128}\); alternatively it might be a problem of model specification instead, more specifically over-fitting the model\(^{129}\). The remedies of such a case could be to omit insignificance variables from the model. This can be ascertained through t-tests and F tests.

The inclusion of irrelevant variables as such does not harm the model. The estimated variances of the coefficients will be larger than the true model resulting in potentially less precise parameters, however the estimates will still be unbiased and consistent and the hypothesis tests so far are still valid\(^{130}\). Over-fitting the model however may lead to multicollinearity\(^{131}\), which is my suspicion of model 2.

Let us refresh the memory of the most noteworthy elements of the estimation of model 2:

\[
\Delta TBI_t = 0.000664 - 0.1040\Delta UNEMP_t - 0.007546\Delta UNEMP_{t-1} - 0.008858\Delta CPI_t - 0.004875\Delta CPI_{t-1} - 0.0688\Delta INT_t + 0.009547\Delta INT_{t-1} + 0.2448\Delta GDP_t + 1.3453\Delta GDP_{t-1} + 0.0519\epsilon_{t-1}
\]

\[
t = \begin{pmatrix}
(0.13) & (-2.52) & (-0.18) & (-0.56) \\
(-0.32) & (-1.82) & (0.26) & (0.37) \\
(1.92) & (2.08) & & 
\end{pmatrix}
\]

\[
R^2 = 0.2227 \quad F = 2.5785 \quad d = 1.8711
\]

Given the satisfactory \(R^2\), significant overall explanatory power given by the F-test and the acceptable value of the Dubin-Watson d statistic the model looks approvable. Assuming a 10% significance level \(\Delta UNEMP_t, \Delta INT_t, \Delta GDP_{t-1}\) and the co-integration error term is appropriate for the model. The inflation given by CPI has some economic intuition relevant for the conclusion of the analysis, thus the insignificance of CPI will not further assessed at

\(^{128}\) For multicollinearity tests see appendix 8 page 90-101. It is evident from the test statistic and graphical view that multicollinearity only exists among the two explanatory variables interest rate and GDP. None of the other variables have any problematic collinearity relationship among them

\(^{129}\) Gujarati, 2003, page 515

\(^{130}\) Gujarati, 2003, page 514

\(^{131}\) Gujarati, 2003, page 514
this point. Nevertheless the amount of statistically insignificant estimates signalize the model could improve.

10.1 Omitting explanatory variables

Given the almost exact linear relationship between interest rate and GDP it is my anticipation at least these variables should be taken into consideration. In addition hereto I am quite skeptical towards the one-period lagged unemployment rate due to its very insignificant test statistic. Omitting variables from a model is a sensitive topic in econometrics, and according to Kerry Patterson (2000)\textsuperscript{132} an interaction between theory and empirical specification is important. In addition Michael Intriligator (1978)\textsuperscript{133} claims explanatory variables should only be included in a model, if they directly influence the dependent variable and if they are not accounted for by other included variables. Thus, my economic intuition anticipate the correlation problem of interest rate and GDP will wane if omitting the insignificant interest rate variable at time $t-1$ and omitting the insignificant GDP variable at time $t$, i.e. keeping a single time series of both (relevant) variables in the model, but influencing the dependent variable at different lags. A joint F-test will clarify the relevance of the variables in the model\textsuperscript{134}. Also, unemployment at time $t-1$ and the insignificant intercept term will be assessed in the F-test.

The model will be estimated under the restricted least squares (RLS) and the F-test will test the following joint hypothesis\textsuperscript{135}:

$$H_0: \beta_1 = \beta_3 = \beta_7 = \beta_8 = 0$$

$$H_1: \text{not } H_0$$

The F-test is:

\textsuperscript{132} Gujarati, 2003, page 517
\textsuperscript{133} Gujarati, 2003, page 514
\textsuperscript{134} For omitting variables’ tests see appendix 9 page 120
\textsuperscript{135} Gujarati, 2003, page 268
From the restricted least squares estimate it is apparent that we cannot reject the joint hypothesis including the four parameters are zero as the F statistic is below the relevant critical value of about 2.70\(^{136}\) i.e. they add no explanatory power to the model and might as well be omitted from the model.

The improved nested model 3 will thus be:

\[
\text{ΔTBI}_t = \beta_1 \text{ΔUNEMP}_t + \beta_2 \text{ΔCPI}_t + \beta_3 \text{ΔCPI}_{t-1} + \beta_4 \text{ΔINT}_t + \beta_5 \text{ΔGDP}_{t-1} + λε_{t-1} + \mu_t
\]

Where \(ΔX_t = \log X_t - \log X_{t-1}\)

En extract of the OLS coefficient estimates of model 3 are\(^{137}\):

\[
\begin{align*}
\Delta \hat{\text{TBI}}_t &= -0.1005 \Delta \hat{\text{UNEMP}}_t - 0.007878 \Delta \hat{\text{CPI}}_t - 0.004950 \Delta \hat{\text{CPI}}_{t-1} \\
&- 0.00771 \Delta \hat{\text{INT}}_t + 1.6614 \Delta \hat{\text{GDP}}_{t-1} + 0.049 \hat{ε}_{t-1} \\
t &= \begin{pmatrix} -3.21 \\ -0.52 \\ -0.34 \end{pmatrix} \quad \begin{pmatrix} -2.31 \\ 8.60 \\ 2.20 \end{pmatrix}
\]

\[R^2 = 0.5284 \quad F = 9.9594 \quad d = 1.8403\]

This nested model is anticipated to be the optimal model for explaining the change in TBI return. From the \(R^2\) and overall significance level \(F\) it is evident that the test did improve our multivariate model quite much, but there are some interesting yet expected features to this model estimation. The nested model highlights the relevance of unemployment, interest rate and GDP at different lags plus the relevance of the co-integration error correction term, while inflation is highly insignificant to the model. As mentioned in the beginning of the paper real estate investments are often used as an inflation hedge, as the income of the

\(^{136}\) The F(m,n-k) lies within \([2.68;2.76]\) given \(n = 120\) or \(n = 60\)

\(^{137}\) For complete outlook on model estimation, please see appendix 9 page 112-115
return will increase by at least equal rate as the inflation rate. Thus the change in inflation ought to be fully covered by the income component of the real estate return rate, and consequently not have any significant power to the change in TBI at any period of time. The hedging capabilities of real estate will be demonstrated next by testing if the two inflation coefficients simultaneously are zero. The anticipation is for the null-hypothesis to be accepted i.e. confirming the inflation-hedging capability of real estate.

From appendix 9 it is apparent that the null-hypothesis of simultaneously zero coefficients of the inflation in time t and t-1 cannot be rejected\(^{138}\); hence inflation at any lags are irrelevant for the model as anticipated.

### 10.2 Estimation of model

Given the expected insignificance of the inflation coefficients, the best obtainable model will then be the following model 4:

\[
\Delta TBI_t = \beta_1 \Delta UNEMP_t + \beta_2 \Delta INT_t + \beta_6 \Delta GDP_{t-1} + \lambda \epsilon_{t-1} + \mu_t
\]

Where \( \Delta X_t = \log X_t - \log X_{t-1} \)

The current changes in TBI is a function of current changes in unemployment, current changes in interest rate and the previous quarter changes in economic growth. Hereto the degree to which the series is outside its’ equilibrium in the previous time period. Note the error correction term will now exclude the inflation factor. The estimated coefficients of model 4 are thus\(^{139}\):

\[
\hat{\Delta TBI}_t = -0.0992 \hat{\Delta UNEMP}_t - 0.0796 \hat{\Delta INT}_t + 1.6564 \hat{\Delta GDP}_{t-1} + 0.0482 \hat{\epsilon}_{t-1}
\]

<table>
<thead>
<tr>
<th>SE</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.0310)</td>
<td>(-3.20)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>(0.0329)</td>
<td>(-2.42)</td>
<td>(0.0175)</td>
</tr>
<tr>
<td>(0.1913)</td>
<td>(8.66)</td>
<td>(&lt;0.0001)</td>
</tr>
<tr>
<td>(0.0215)</td>
<td>(2.24)</td>
<td>(0.0278)</td>
</tr>
</tbody>
</table>

\(^{138}\) See appendix 9 page 120 for F-test statistics on \( H_0: \beta_2 = \beta_3 = 0 \)

\(^{139}\) See appendix 9 page 116-119
The model goodness of fit has heavily improved from model 2, mainly due to removal of the intercept term. This is not in itself a target; however given the normality of the regression model, an acceptable value of the Durbin-Watson statistic, and the improved value of the Akaike Information Criterion indicate the development of a better model. The significance of all the explanatory variables further substantiates the analysis of the macroeconomic time series variables’ influence on the US real estate transaction-based return.

10.3 The macroeconomic impact on real estate returns

With respect to the estimate of model 4, it can be concluded that changes in US real estate return as expected is significantly negatively affected by an increase in the current unemployment rate and equally by the change in current interest rate. Oppositely the economic growth of United States in the previous quarter given by GDP has a very significant positive influence on changes in TBI. These results all correspond to the correlation coefficients among the macroeconomic variables and the real estate index. As anticipated the inflation change has no influence on the return index, however this does not necessarily imply that either the income component (or the capital appreciation component) individually is not influenced, simply that on an aggregate level inflation changes is not captured in the TBI change over time. The small significant error correction term assures long term equilibrium among the explanatory variables that co-trends stably over time. The significance level of the error term further proves the co-integration properties of the variables.
Chapter 11
Conclusion

In this paper I have analysed the macroeconomic influence on the United States real estate return. I have found that changes in the macro economy as anticipated do significantly affect the changes in the transaction-based real estate return. The analysis was performed through a log-linear regression model based on academic research by De Wit and Van Dijk.

The thesis began with an introduction to the real estate market and highlighted its limitations in terms of an illiquid and non-transparent market place. Conversely real estate has numerous qualities, in particular the inflation hedge capabilities diverse real estate from other investment assets.

I found that valuation of properties is incredibly complicated due to the associated statistical errors of the valuation methods. The first approach given by the appraisal based valuation suffers from temporal lag bias as smoothing processes of the valuation lead to incorrectly reduced volatility and bias the return to the preceding periodic trend. The second approach given by transaction price based valuation suffers from pure random error effects. The random errors cause spurious error terms that overestimate the volatility and reduces the potential correlation to exogenous series. A trade-off between the errors makes it difficult to reduce either without increasing the other, however unsmoothing of the appraisal-based returns is found to be the best adjusting method.
Three methods of unsmoothing was highlighted, the first being a zero-autocorrelation technique that statistically remove autocorrelation from the residuals through a manually chosen constant factor. The second method is a mechanical de-lagging approach that inverse a weighted first-order autoregressive model to obtain the transaction-based return. Both methods do improve the index and increase the volatility, however, the constant factor and lag-weight, respectively, are theoretically chosen and the models are not very dynamic. The third and final model is an econometric approach that unsmoothes the return series through a hedonic regression model. The hedonic log-function typically contains more than 60 explanatory variables, and is the best estimation of the true market returns. The TBI return uses such a hedonic approach and was used in the empirical analysis.

The empirical analysis resulted in a number of interesting conclusions. I found the I(1) stationary macroeconomic variables to have a co-integrated relationship, thus the estimation was based on an error correction model including the macro variables in their level and lagged form plus a long-term equilibrium error term. The final model fulfilling the assumptions of the classical linear regression model was a function of changes in log values of the transaction based return regressed on unemployment and interest rate at time t and GDP one period lagged. As anticipated the inflation given by CPI was not significant due to real estate returns qualities as an inflation hedge. Thus, inflation is assumed to be given by the income component of the return and therefore has no influence on the changes in the total return.

In respect to the problem statement, the estimation of the model showed that unemployment has a significant negative influence on changes in the return, and same was the result for the interest rate. GDP in the previous quarter oppositely had a significant influence on the change in return. This makes good economical sense as real estate intuitively is regarded a good indicator of the general economy and mirror its’ development. The estimates for inflation were negative as expected from the problem statement hypothesis, but not statistically significant to the model. Lastly, the significant positive long term equilibrium error term, given by the co-integration among the
exploratory variables, showed that the transaction based return is below its equilibrium level and the error term will cause the return to increase in the following period.

The findings of the United States total return model mirrors surprisingly well the results of the De Wit and Van Dijk global office analysis I based my model upon. The global office model likewise resulted in a significantly positive GDP, and equivalent significant negative unemployment rate, while inflation was insignificant on a total return level. This opposes some of the other academic literature on the topic; however, analyzing the exact same macroeconomic variables seemingly have the same result on a US cross-sector return level as that of the global office market.

**In conclusion** my analysis demonstrates that market fundamentals given by changes in the macro economy over a period of twenty-five years do significantly influence the United States real estate return level. The model estimation demonstrated more than 52% of the change in return can be explained by the macroeconomic variables in question; accordingly, property investment is not only a question of managerial skills within real estate but is also highly dependent on market timing and forecasting skills within macroeconomics.
Chapter 12
Perspective

In this paper I have demonstrated the importance of macroeconomic development to real estate returns in the United States. It is inevitably important for real estate investors to be aware of the huge influence real market forces have on the return of their investments, and be knowledgeable about the market timing perspective of their investments.

In the delimitation of the paper, it was my choice solely to focus on the total return base in the US market. However, the structure of the regression model developed here can easily be used as a framework for other purposes. What could be of interest would be a separate analysis of the macroeconomic determinants on each component of the return given by income and capital appreciation. An empirical investigation of such could potentially lead to different results on a disaggregate level and further clarify the actual inflation hedge capabilities of the income return.

Moreover, other explanatory variables could be of interest in such an analysis, being both other macroeconomic variables and more real estate specific variables. De Wit and Van Dijk also refer to a number of office-specific variables as e.g. vacancy rate and supply of office stock that could easily be applied to a more overall real estate return model. Lastly an analysis of the different sub-segments or geographical differences to the real estate market return could add value for acquisition managers within real estate. However, in my view it would demand a more transparent market e.g. in terms of availability of sub-segment return index from the NPI database.
By writing this paper, however, it is my anticipation the analysis did add value to the overall understanding of the interdependence between the real estate market and the fundamental macro economy.
Chapter 13

Literature List

13.1 Books and Articles

Andersen & Hjortshøj, Finans/Invest Vol. 5, 2008
Marked for ejendomsderivater I Danmark

Benninga, Simon, MIT Press
Financial Modeling, 2nd edition

Macroeconomics, 2nd edition

Real Estate Returns and Inflation: An Added Variables Approach

Brealey, R., Myers, S., and Allen, F., 2006, McGraw-Hill/Irwin
Corporate Finance, 8th edition

The Econometrics of Financial Markets page 188

Case, Goetzmann and Rouwenhorst. Working paper, NBER, 2000
Global Real Estate Markets – Cycles and Fundamentals.

Clayton, Jim, PREA – Pension Real Estate Association, October 1, 2008
Capital Market Overview

De Wit, I., and Van Dijk, R, 2003, Journal of Real Estate Finance & Economics,
26:1, 27-45. The Global Determinants of Direct Office Real Estate Return

The Determinants of Commercial Property Prices and Rents

Elton, E., Gruber, M., Brown, S. and Goetzmann, W., 2007, John Wiley and Sons, Ltd
Macroeconomic Determinants of Real Estate Returns

*Modern Portfolio Theory and Investment Analysis*

Ernst & Young, 2007
*Market Outlook – Trends in the Real Estate Private Equity Industry*

Fisher, Geltner, Pollakowski, 2006, MIT Center of Real Estate
*A Quarterly Transaction-Based Index (TBI) of Institutional Real Estate Investment Performance and Movements in Supply and Demand*

Geltner, Miller, Clayton and Eicholtz, 2006, Thompson
*Commercial Real Estate Analysis & Investments, 2nd edition*

*Benefits of Real Estate Investment*

*Basic Econometrics, 4th edition*

*Real Estate Returns and Inflation*

*Rental Price Adjustment and Investment in the Office Market*

*Private Commercial Real Estate Equity Returns and Inflation*

*Why Real Estate?*

*Hedonic Price Models and Indices Based on Boosting Applied to the Dutch Housing Market page 91*

Koop, Gary 2008, John Wiley & Sons, Ltd
*Introduction to Econometrics*

*Economic Risk Factors and Commercial Real Estate Returns*

Marcato and Key, 2007, the Journal of Portfolio Management
*Smoothing and implications for asset allocation choices*

Page 88 of 113
McIntosh, A., 1989, Pensions, August, 35-36
Building on firm foundations

Peyton, M., Park, T., and Badillo, F., 2006, TIIA-CREF asset management
Why Real Estate?

Private Equity Real Estate, May 2008, vol 4, issue 4 page 50

Stevenson, Kinsella and O’Healai, 1997, Irish Business & Administrative Research 18, 163-76
Irish Property Funds: Empirical Evidence on Market Timing and Selectivity

The Journal of Portfolio Management, September 2003

The Impact of Inflation and Vacancy of Real Estate Returns

Young, Michael, 2005, Journal of Real Estate Portfolio Management, Vol. 11, No. 3
Making sense of the NCREIF Property Index: A New Formulation Revisited

Introduction to Time Series Analysis and Forecasting.

13.2 Websites

Danske Analyse, Danske Bank
www.danskeanalyse.danskebank.dk

ING
www.ing.com

ING Clarion
www.ingclarion.com

Investment Property Databank
www.ipd.com

Massachusetts Institute of Technology - MIT Centre for Real Estate
http://web.mit.edu/cre/

National Council of Real Estate Investment Fiduciaries
www.ncreif.com

The Journal of Portfolio Management via CBS library EBSCO host
13.3 Other

Invesco, 2008, IPE European Institutional Asset Management Survey 2008 page 9-10

La Cour, 2006, Applied Econometrics lecture notes

SAS Institute A/S, the SAS System Version 9.1.3: Enterprise Guide

Thomson Financials, Datastream Advance 4.0
## Table of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix 1</td>
<td>TIME SERIES VARIABLES</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>UNSMOOTHING TECHNIQUES</td>
</tr>
<tr>
<td>Appendix 3</td>
<td>STATIONARITY</td>
</tr>
<tr>
<td>Appendix 4</td>
<td>CO-INTEGRATION</td>
</tr>
<tr>
<td>Appendix 5</td>
<td>NORMALITY</td>
</tr>
<tr>
<td>Appendix 6</td>
<td>ARCH</td>
</tr>
<tr>
<td>Appendix 7</td>
<td>AUTOCORRELATION</td>
</tr>
<tr>
<td>Appendix 8</td>
<td>MULTICOLLINEARITY</td>
</tr>
<tr>
<td>Appendix 9</td>
<td>MODEL ESTIMATION</td>
</tr>
</tbody>
</table>

A greater part of the appendices are output from SAS, and are only disclosed to the printed version of the thesis. Please contact supervisor Lisbeth La Cour or the writer of the thesis Thomas Kofoed-Pihl for further information on the appendices.
Appendix 1

Time Series Variables

The appendix contains the time series variables used in the log-linear regression model of the paper. In addition hereto an outlook of the series over time, and covariance/correlation coefficients among the variables are disclosed.

- Time series variables
  - TBI
  - Unemployment
  - CPI
  - Interest Rate
  - GDP
- Time series graphs
- Covariances / Correlation coefficients
Macroeconomic Determinants of Real Estate Returns

Delta Log Interest Rate

Log Interest Rate

Delta Log GDP

Log GDP
### Covariances (log values)

<table>
<thead>
<tr>
<th></th>
<th>TBI</th>
<th>UNEMP</th>
<th>CPI</th>
<th>INT</th>
<th>GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBI</td>
<td>0,076908</td>
<td>-0,00384</td>
<td>-0,01134</td>
<td>-0,0325</td>
<td>0,043318</td>
</tr>
<tr>
<td>UNEMP</td>
<td>0,004075</td>
<td>0,000132</td>
<td>0,000395</td>
<td>-0,00206</td>
<td></td>
</tr>
<tr>
<td>CPI</td>
<td>0,024554</td>
<td>0,009258</td>
<td>0,0191</td>
<td>-0,02079</td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,027265</td>
</tr>
<tr>
<td>GDP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Correlations (log values)

<table>
<thead>
<tr>
<th></th>
<th>TBI</th>
<th>UNEMP</th>
<th>CPI</th>
<th>INT</th>
<th>GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBI</td>
<td>1</td>
<td>-0,21841</td>
<td>-0,26312</td>
<td>-0,8525</td>
<td>0,954163</td>
</tr>
<tr>
<td>UNEMP</td>
<td>1</td>
<td>0,013174</td>
<td>0,044752</td>
<td>-0,19564</td>
<td></td>
</tr>
<tr>
<td>CPI</td>
<td>1</td>
<td>0,427518</td>
<td>0,427518</td>
<td>-0,28119</td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>-0,91084</td>
</tr>
<tr>
<td>GDP</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Correlations (delta log values)

<table>
<thead>
<tr>
<th></th>
<th>TBI</th>
<th>UNEMP</th>
<th>CPI</th>
<th>INT</th>
<th>GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBI</td>
<td>1</td>
<td>-0,22273</td>
<td>-0,08187</td>
<td>-0,07572</td>
<td>0,004908</td>
</tr>
<tr>
<td>UNEMP</td>
<td>1</td>
<td>-0,04444</td>
<td>-0,07399</td>
<td>-0,23027</td>
<td></td>
</tr>
<tr>
<td>CPI</td>
<td>1</td>
<td>0,109289</td>
<td>0,087028</td>
<td>0,123071</td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2

Unsmoothing techniques

The appendix contains a sample of potential variables to be included in a hedonic real estate unsmoothing regression model. The unsmoothing techniques of Zero-Autocorrelation and the Mechanical De-lagging approach are available from CD-Rom

- Zero-autocorrelation technique (Available from CD-Rom)
- Mechanical de-lagging (Available from CD-Rom)
- Hedonic Regression Model (Disclosed)
### Hedonic Regression Model Variables - Kagie and Wezel, 2008.
*John Wiley & Sons Ltd. Hedonic Price Models and Indices*

#### Table 1: List of variables included in the data set. Variables 1-12 are categorical variables (number of levels between brackets).

<table>
<thead>
<tr>
<th>Variable Code</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>Dwelling size (m²)</td>
</tr>
<tr>
<td>23</td>
<td>Main house (status)</td>
</tr>
<tr>
<td>24</td>
<td>Structure (status)</td>
</tr>
<tr>
<td>25</td>
<td>Room availability</td>
</tr>
<tr>
<td>26</td>
<td>Type of building</td>
</tr>
<tr>
<td>27</td>
<td>Building condition</td>
</tr>
<tr>
<td>28</td>
<td>Building size (m²)</td>
</tr>
<tr>
<td>29</td>
<td>Building materials</td>
</tr>
<tr>
<td>30</td>
<td>Building location</td>
</tr>
<tr>
<td>31</td>
<td>Building orientation</td>
</tr>
<tr>
<td>32</td>
<td>Building condition</td>
</tr>
<tr>
<td>33</td>
<td>Building size (m²)</td>
</tr>
<tr>
<td>34</td>
<td>Building materials</td>
</tr>
<tr>
<td>35</td>
<td>Building location</td>
</tr>
<tr>
<td>36</td>
<td>Building orientation</td>
</tr>
<tr>
<td>37</td>
<td>Building condition</td>
</tr>
<tr>
<td>38</td>
<td>Building size (m²)</td>
</tr>
<tr>
<td>39</td>
<td>Building materials</td>
</tr>
<tr>
<td>40</td>
<td>Building location</td>
</tr>
<tr>
<td>41</td>
<td>Building orientation</td>
</tr>
<tr>
<td>42</td>
<td>Building condition</td>
</tr>
<tr>
<td>43</td>
<td>Room availability (no. of rooms)</td>
</tr>
<tr>
<td>44</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>45</td>
<td>Room condition</td>
</tr>
<tr>
<td>46</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>47</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>48</td>
<td>Room condition</td>
</tr>
<tr>
<td>49</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>50</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>51</td>
<td>Room condition</td>
</tr>
<tr>
<td>52</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>53</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>54</td>
<td>Room condition</td>
</tr>
<tr>
<td>55</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>56</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>57</td>
<td>Room condition</td>
</tr>
<tr>
<td>58</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>59</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>60</td>
<td>Room condition</td>
</tr>
<tr>
<td>61</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>62</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>63</td>
<td>Room condition</td>
</tr>
<tr>
<td>64</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>65</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>66</td>
<td>Room condition</td>
</tr>
<tr>
<td>67</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>68</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>69</td>
<td>Room condition</td>
</tr>
<tr>
<td>70</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>71</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>72</td>
<td>Room condition</td>
</tr>
<tr>
<td>73</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>74</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>75</td>
<td>Room condition</td>
</tr>
<tr>
<td>76</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>77</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>78</td>
<td>Room condition</td>
</tr>
<tr>
<td>79</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>80</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>81</td>
<td>Room condition</td>
</tr>
<tr>
<td>82</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>83</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>84</td>
<td>Room condition</td>
</tr>
<tr>
<td>85</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>86</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>87</td>
<td>Room condition</td>
</tr>
<tr>
<td>88</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>89</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>90</td>
<td>Room condition</td>
</tr>
<tr>
<td>91</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>92</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>93</td>
<td>Room condition</td>
</tr>
<tr>
<td>94</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>95</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>96</td>
<td>Room condition</td>
</tr>
<tr>
<td>97</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>98</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>99</td>
<td>Room condition</td>
</tr>
<tr>
<td>100</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>101</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>102</td>
<td>Room condition</td>
</tr>
<tr>
<td>103</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>104</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>105</td>
<td>Room condition</td>
</tr>
<tr>
<td>106</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>107</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>108</td>
<td>Room condition</td>
</tr>
<tr>
<td>109</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>110</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>111</td>
<td>Room condition</td>
</tr>
<tr>
<td>112</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>113</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>114</td>
<td>Room condition</td>
</tr>
<tr>
<td>115</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>116</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>117</td>
<td>Room condition</td>
</tr>
<tr>
<td>118</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>119</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>120</td>
<td>Room condition</td>
</tr>
<tr>
<td>121</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>122</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>123</td>
<td>Room condition</td>
</tr>
<tr>
<td>124</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>125</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>126</td>
<td>Room condition</td>
</tr>
<tr>
<td>127</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>128</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>129</td>
<td>Room condition</td>
</tr>
<tr>
<td>130</td>
<td>Room accessibility</td>
</tr>
<tr>
<td>131</td>
<td>Room size (m²)</td>
</tr>
<tr>
<td>132</td>
<td>Room condition</td>
</tr>
</tbody>
</table>
Appendix 3

Stationarity

The appendix contains stationarity tests of the individual time series variables.

The disclosed test statistics are:

- Durbin-Watson d-statistic
- The Breusch-Godfrey test (Lagrange Multiplier test)
- ACF / PACF correlograms
- The (augmented) Dickey-Fuller Unit Root Tests
Appendix 4

Co-integration

The appendix contains co-integration tests among the time series variables. The tests are performed on the original log-linear regression model, on a model including all the explanatory variables, and pair-wise of the explanatory macroeconomic variables.

The disclosed test statistics are:

- Co-integrating Regression Durbin-Watson Test
- Augmented Engle-Granger Test
Appendix 5

Normality

The appendix contains test of normal distribution of the error correction model 1.

\[
\Delta TBI_t = \beta_1 + \beta_2 \Delta UNEMP_t + \beta_3 \Delta UNEMP_{t-1} + \beta_4 \Delta CPI_t + \beta_5 \Delta CPI_{t-1} + \\
\beta_6 \Delta INT_t + \beta_7 \Delta INT_{t-1} + \beta_8 \Delta GDP_t + \beta_9 \Delta GDP_{t-1} + \lambda_2 e_{t-1} + \mu_t
\]

In addition hereto the normality tests are performed on the three adjusted error correction models, subject to paragraph 7.3.2 of the thesis.

The disclosed test statistics are:

- Histogram
- Probability plots
- Jarque-Bera Normality Test
- Ramsey’s Regression Specification Error Test
- Model selection through Dummy Regression Analysis
**Error Correction Model Selection – Dummy Regression**

In order to correct for outliers in the error correction model, a *dummy variable of 1* is included in the regression model at time $t$ of the outlier. To assure immediate adjustment for the delta lag values a *dummy variable of -1* is included at time $t+1$ of the outlier.

Dummy:

- 1 if outlier
- -1 at $t+1$ of outlier

<table>
<thead>
<tr>
<th>Model selection</th>
<th>Original ECM 1</th>
<th>Dummy Adj ECM 2</th>
<th>Dummy Adj ECM 3</th>
<th>Dummy Adj ECM 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Year; Quarter)</td>
<td></td>
<td>-</td>
<td>1988:4</td>
<td>1988:4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1991:4</td>
<td>1991:4</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2007:4</td>
<td>-</td>
<td>2007:4</td>
</tr>
<tr>
<td>AIC</td>
<td>-514.21094</td>
<td>-522.07626</td>
<td>-521.98555</td>
<td>-521.62251</td>
</tr>
<tr>
<td>SIC</td>
<td>-488.56746</td>
<td>-493.86843</td>
<td>-493.77772</td>
<td>-493.41468</td>
</tr>
</tbody>
</table>
Appendix 6

ARCH

The ARCH tests are performed on an auxiliary regression of model 2 the estimated squared residuals at different lags

\[ \hat{\mu}_t^2 = \gamma_0 + \gamma_1 \hat{\mu}_{t-1}^2 + \ldots + \gamma_p \hat{\mu}_{t-p}^2 + \text{error}_t \]

The disclosed test statistics are:

- ACF / PACF
- Engle’s ARCH test based on an auxiliary regression
- Engle’s test in SAS (LM test for ARCH disturbances)
Appendix 7

Autocorrelation – model 2

The appendix contains of residual plots of model 2. The two numerical autocorrelations tests performed can be seen from the estimation of the error correction model 2 in appendix 5.

The disclosed test statistics are:

- Breush-Godfrey test (See ECM 2 from appendix 5)
- Durbin-Watson test (See ECM 2 from appendix 5)
- Residuals plots
Appendix 8

Multicollinearity

The appendix contains investigation of the pair-wise linear relationship between the explanatory macroeconomic variables. Disclosed are:

- Scatter plots
- Linear relationship between the explanatory variables
- Auxiliary F-tests
Auxiliary F-tests

\[ F = \frac{R^2/(k-2)}{(1-R^2)/(n-k+1)} \]

**Auxiliary regression: \( \log(\text{UNEMP}) = b_1 + b_2 \log(\text{CPI}) + b_3 \log(\text{INT}) + b_4 \log(\text{GDP}) \)**

<table>
<thead>
<tr>
<th></th>
<th>R2</th>
<th>k</th>
<th>n</th>
<th>F</th>
<th>Critical value 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1475</td>
<td>4</td>
<td>93</td>
<td>7.785924</td>
<td>60df: 2.76, 120df: 2.68</td>
</tr>
<tr>
<td></td>
<td>0.07375</td>
<td>0.009472</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F-test: all coefficients equal zero
Reject that model equals zero, it is significantly different from zero

**Auxiliary regression: \( \log(\text{CPI}) = b_1 + b_2 \log(\text{UNEMP}) + b_3 \log(\text{INT}) + b_4 \log(\text{GDP}) \)**

<table>
<thead>
<tr>
<th></th>
<th>R2</th>
<th>k</th>
<th>n</th>
<th>F</th>
<th>Critical value 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.2595</td>
<td>4</td>
<td>93</td>
<td>15.76975</td>
<td>60df: 2.76, 120df: 2.68</td>
</tr>
<tr>
<td></td>
<td>0.12975</td>
<td>0.008228</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F-test: all coefficients equal zero
Reject that model equals zero, it is significantly different from zero

**Auxiliary regression: \( \log(\text{INT}) = b_1 + b_2 \log(\text{UNEMP}) + b_3 \log(\text{CPI}) + b_4 \log(\text{GDP}) \)**

<table>
<thead>
<tr>
<th></th>
<th>R2</th>
<th>k</th>
<th>n</th>
<th>F</th>
<th>Critical value 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.8738</td>
<td>4</td>
<td>93</td>
<td>311.5769</td>
<td>60df: 2.76, 120df: 2.68</td>
</tr>
<tr>
<td></td>
<td>0.4369</td>
<td>0.001402</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F-test: all coefficients equal zero
Reject that model equals zero, it is significantly different from zero
Auxiliary regression: $\log(\text{GDP}) = b_1 + b_2 \log(\text{UNEMP}) + b_3 \log(\text{CPI}) + b_4 \log(\text{INT})$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td>0.8631</td>
<td>0.43155</td>
</tr>
<tr>
<td>k</td>
<td>4</td>
<td>0.001521</td>
</tr>
<tr>
<td>n</td>
<td>93</td>
<td></td>
</tr>
</tbody>
</table>

$F = 283.7071$

Critical value 5%  
60df  
120df  

2.76  
2.68

F-test: all coefficients equals zero
Reject that model equals zero, it is significantly different from zero
Appendix 9

Model Estimation

The appendix contains the estimation of the four models developed through the empirical analysis. In addition hereto overall significance tests of the models are disclosed. The estimated models disclosed are:

- Model 1 – Error Correction Model
- Model 2 – Adjusted Error Correction Model
- Model 3
- Model 4
- Omitting Variables: model 3 and model 4 (F-tests)
- Overall significance test of models (F-test)
Error correction model 1
Error correction model 2
Model 3
Model 4
**Overall Significance test – F-tests**

\[ F = \frac{R^2 / (k-1)}{(1 - R^2) / (n-k)} \sim F(k-1, n-k) \]

<table>
<thead>
<tr>
<th>Original ECM model</th>
<th>R2</th>
<th>0.126</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>96</td>
<td></td>
</tr>
</tbody>
</table>

\[ F = 1.377574 \]

Critical value 5% 60df 2.04

120df 1.96

Do not reject that model equals zero

<table>
<thead>
<tr>
<th>Adjusted ECM model 1</th>
<th>R2</th>
<th>0.2227</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>91</td>
<td></td>
</tr>
</tbody>
</table>

\[ F = 2.578541 \]

Critical value 5% 60df 2.04

120df 1.96

Reject that model equals zero, it is significantly different from zero

<table>
<thead>
<tr>
<th>Nested model 1</th>
<th>R2</th>
<th>0.5284</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>90</td>
<td></td>
</tr>
</tbody>
</table>

\[ F = 9.959476 \]

Critical value 5% 60df 2.04

120df 1.96

Reject that model equals zero, it is significantly different from zero

<table>
<thead>
<tr>
<th>Best model</th>
<th>R2</th>
<th>0.5259</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>90</td>
<td></td>
</tr>
</tbody>
</table>

\[ F = 9.860086 \]

Critical value 5% 60df 2.04

120df 1.96

Reject that model equals zero, it is significantly different from zero
Omitting Variables – F-tests

\[ F = \frac{R^2_{\text{ur}} - R^2_{\text{r}}}{m} \frac{1 - R^2_{\text{ur}}}{n - k} \]

Model 3
Omit: Intercept, Delta Lag UNEMP, Delta Lag INT and Delta GDP

Ho: \( B_1 = B_3 = B_7 = B_8 = 0 \)

\[
R^2 (r) = 0.2197 \\
R^2(\text{ur}) = 0.2227
\]

\[
m = 4 \\
k = 10 \\
n = 90
\]

\[
F = \frac{(R^2(\text{ur})-R^2(r))/m}{(1-R^2(\text{ur}))/(n-k)} \\
F = \boxed{0.07719}
\]

F(0.95)(m,n-k) = F(4,80) = 2.53

Accept null-hypothesis of no explanatory power

Model 4
Omit: Lag CPI and Delta CPI

Ho: \( B_4 = B_5 = 0 \)

\[
R^2(r) = 0.5259 \\
R^2(\text{ur}) = 0.5284
\]

\[
m = 2 \\
k = 10 \\
n = 90
\]

\[
F = \frac{(R^2(\text{ur})-R^2(r))/m}{(1-R^2(\text{ur}))/(n-k)} \\
F = \boxed{0.212044}
\]

F(0.95)(m,n-k) = F(2,80) = 3.15

Accept null-hypothesis of no explanatory power