Valuation of Properties and Economic Models of Real Estate Markets

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Dr. rer. pol.)
im Fach Ökonometrie und Volkswirtschaftslehre

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von
Diplom-Volkswirt M.A. Rainer Schulz
geboren am 10.01.1967 in Jugenheim an der Bergstraße

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Michael C. Burda, Ph.D.

Gutachter:

1. Prof. Dr. Wolfgang Härdle
2. Prof. Irwin L. Collier, Ph.D.

eingereicht am:	3. Januar 2003
Tag der mündlichen Prüfung:	5. Februar 2003

Abstract

Appraisals should assess the market value of properties and are necessary for buying, selling or building decisions, for lending and for taxation. Despite this unambiguous task different techniques exist for ascertaining market values. An valuation approach should be in accordance with economic theory and should generate appraisals, which are reliable estimates for transaction prices. This dissertation analyzes the three most important valuation approaches, i.e. cost, sales comparison, and income approach, shows the underlying market models and evaluates the valuation techniques that are codified in the German Regulation on Valuation (WertV). For the latter evaluations, appraisals are compared with observed transaction prices. In addition, the dissertation gives an overview on real estate price indices and on the hedonic approach. Extensive data on Berlin's real estate market are used for the econometric analysis.

Keywords:
Real estate price indices, Hedonic approach, Appraisal accuracy, State space model

Zusammenfassung

Bewertungen von Immobilien sollen den Marktwert einschätzen und sind notwendig für Kauf-, Verkaufs- und Bauentscheidungen, für die Kreditvergabe und für die Besteuerung. Trotz dieser eindeutigen Aufgabenstellung existierten unterschiedliche Verfahren, mit welchen Marktwerte ermittelt werden können. Ein Bewertungsverfahren soll einerseits mit ökonomischer Theorie vereinbar sein und andererseits Bewertungen generieren, die beobachtete Transaktionspreise gut vorhersagen. Die Dissertation analysiert die drei wichtigsten Bewertungsansätze Sachwert-, Vergleichswert- und Ertragswertverfahren, zeigt das jeweils zugrundeliegende Marktmodell und evaluiert die kodifizierten Verfahren nach der Wertermittlungsverordnung (WertV) anhand von beobachteten Transaktionen. Darüber hinaus gibt die Dissertation einen Überblick zu Immobilienpreisindizes und zu hedonischen Methoden. Für die ökonometrischen Analysen wurden umfangreiche Daten zum Berliner Immobilienmarkt verwendet.

Schlagwörter:
Immobilien-Preisindexe, Hedonische Methode, Bewertungsgüte, State space Modell

Danksagung

Die vorliegende Arbeit behandelt die Preisbildung und die Bewertung von Immobilien. Mit Hilfe ökonomischer Theorie wird gezeigt, dass Bewertungsverfahren auf plausiblen Marktmodellen beruhen. Dies wird genutzt, um die Verfahren empirisch zu evaluieren.

Die Arbeit hätte ohne das Engagement und die Energie meines Erstgutachters Prof. Dr. Wolfgang Härdle nicht entstehen können. Durch seine Kontakte erhielt ich Zugang zu einzigartigen Daten, mit welchen ich unter optimalen technischen Bedingungen arbeiten konnte. Mit seinem Blick für das Wesentliche hat er mich vor Umwegen bewahrt und stand mir immer mit Rat und Tat zur Seite. Bei meinem Zweitgutachter Prof. Irwin L. Collier möchte ich mich sehr dafür bedanken, dass er mich zu meinem Thema geführt hat und mich bei meiner Arbeit mit hilfreichen Diskussionen und Vorschlägen begleitet hat. Weiterhin möchte ich mich insbesondere bei Alan Kirman, Robert J. Shiller, Stefano G. Athanasoulis und den Mitarbeitern der Geschäftsstelle des Gutachterausschusses für Grundstückswerte in Berlin für hilfreiche Diskussionen und Anregungen bedanken.

Bei allen Kollegen am Institut für Statistik, dem Sonderforschungsbereich 373 und im XploRe Team bedanke ich mich für ihre Unterstützung und die kooperative Arbeitsatmosphäre. Besonderer Dank gilt Axel Werwatz, Rodrigo Witzel und Hizir Sofyan, die mit mir MD*Immo umgesetzt haben, Lenka Cizkova für ihre numerischen Methoden in XploRe, Torsten Kleinow und Nikolaj Schmuklermann. Bei Scott Solano bedanke ich mich für das Korrekturlesen von Teilen dieser Arbeit und bei Corinna Hinke für die Grafiken und das Design von MD*Immo.

Bei der Deutschen Forschungsgemeinschaft bedanke ich mich für die Finanzierung, die mir durch das Graduiertenkolleg *Angewandte Mikroökonomik* und den Sonderforschungsbereich 373 *Quantifikation und Simulation Ökonomischer Prozesse* gewährt wurde.

Contents

Contents iv

Glossary vii

List of Figures xi

List of Tables xii

1 Introduction 1
- 1.1 The importance of real estate appraisals 1
- 1.2 Structure of the study 3

2 Measures of Real Estate Prices 7
- 2.1 Valuation approaches 7
  - 2.1.1 What is a valuation approach? 7
  - 2.1.2 Sales comparison approach 11
  - 2.1.3 Income approach 14
  - 2.1.4 Cost approach 16
  - 2.1.5 Final reconciliation and combinations of the approaches 17
  - 2.1.6 Codified valuation techniques in Germany 17
  - 2.1.7 Which valuation approach should be used? 19
- 2.2 Models for the real estate market 20
  - 2.2.1 Asset market approach 21
  - 2.2.2 Location approach 27
  - 2.2.3 Summary on the rationales of valuation approaches 32
- 2.3 Index numbers for real estate 33
  - 2.3.1 Market indices for assets 34
  - 2.3.2 Difficulties in calculating real estate indices 38

2.3.3 Price indices for real estate in Germany 40
2.3.4 Summary on German real estate indices 44
2.4 Statistical real estate price indices 45
2.4.1 Hedonic approach 45
2.4.2 Repeat sales technique 53
2.4.3 Which technique should be used? 56

3 A State Space Model for Berlin Single-Family House Prices 58
3.1 Introduction 58
3.2 Value of a house 60
3.3 State space model and estimation algorithm 64
3.4 Empirical investigation 71
3.4.1 Data 71
3.4.2 Empirical results 74
3.5 Conclusion 81
3.6 Appendix 82
3.6.1 Kalman filter recursions 82
3.6.2 Reference value 87
3.6.3 Market indicators 89

4 MD*Immo—Online Prediction of Berlin Single-Family House Prices 91
4.1 Introduction 91
4.2 Calculation of the price predictions 92
4.2.1 Time structure and data 92
4.2.2 Price predictions 93
4.3 Technical implementation 96
4.4 Outlook 99

5 An Empirical Analysis of Income Valuation in Germany 102
5.1 Introduction 102
5.2 Income approach 103
5.2.1 Valuation according to WertV 104
5.2.2 Valuation according to the multiple technique 106
5.3 How to evaluate different valuation techniques? 107
5.3.1 Statistical model based on WertV 108
5.3.2 What are the general market conditions? 109
5.4 Empirical investigation 111
5.4.1 Data 111

5.4.2 Assessment of valuations according to WertV 113
5.4.3 Comparison with the multiple technique 117
5.5 General market conditions 121
5.5.1 Market indicators 122
5.6 Conclusion 125
5.7 Appendix 127

6 An Empirical Analysis of Cost Valuation in Germany 128
6.1 Introduction 128
6.2 Cost approach 130
6.3 Empirical investigation 133
6.3.1 Data 133
6.3.2 Empirical results 134
6.4 Conclusion 143
6.5 Appendix 144

Bibliography 145

Glossary

Abbreviations

AKS	*Automatisierte Kaufpreissammlung* Automated Database of Transaction Prices
BaFin	*Bundesanstalt für Finanzdienstleistungsaufsicht* German Financial Supervisory Authority
BauGB	*Baugesetzbuch* German Building Law
BGB	*Bürgerliches Gesetzbuch* German Civil Code
CASE	Center for Applied Statistics and Economics at the Humboldt-Universität zu Berlin
DEIX	*Deutscher Eigentums-Immobilien-Index* German Real Estate Ownership Index
DID	*Deutsche Immobilien Datenbank* German Property Database
DIX	*Deutscher Immobilien Index* German Property Index
EigZulG	*Eigenheimzulagengesetz* Ownership Promotion Law
GAA	*Gutacherausschuss für Grundstückswerte in Berlin* Berlin's Surveyor Commission for Real Estate
GSt	*Geschäftsstelle des GAA* Office of Berlin's Surveyor Commission
HypBankG	*Hypothekenbankgesetz* German Mortgage Banks Act

KAGG	*Gesetz über Kapitalanlagegesellschaften* German Capital Investment Companies Act
RDM	*Ring Deutscher Makler* Federal Organisation of Estate Agents and Property Management e.V.
RICS	Royal Institution of Chartered Surveyors
SSM	State Space Model
StaLa	*Statistisches Landesamt Berlin* Statistical Office Berlin
VDM	*Verband Deutscher Makler* Organization for Property Management and Financing e.V.
WertR 91	*Wertermittlungs-Richtlinien* Guidelines on Valuation
WertV	*Wertermittlungsverordnung* Regulation on Valuation
ZVG	*Gesetz über die Zwangsversteigerung und Zwangsverwaltung* Law on Forced Sales and Forced Administration

Frequently used notation

$x \stackrel{\text{def}}{=} \dots$	x is defined as $\dots$
$A^\top$	transpose of matrix A
$B_{n,t}$	costs for building a structure with the characteristics of property n in period t
$C_{n,t}$	indicated value for property n in period t given by the cost approach
$\mathcal{C}[X,Y]$	covariance of two random variables X and Y
$D_{n,t}$	net operating rent of property n in period t
$E_{n,t}$	indicated value for property n in period t given by the income capitalization approach
$\mathcal{E}[X]$	expected value of random variable X
$\mathcal{E}_t[X]$	conditional expected value of random variable X given information set $\mathcal{F}_t$
$\mathcal{F}_{n,t}$	information set generated by all information available in period t
I_n	$(n \times n)$ identity matrix
$L_{n,t}$	value of the lot for property n in period t
M_t	general market conditions in period t
$N(\mu,\Sigma)$	Normal distribution with expectation μ and covariance matrix Σ
Ω_t	information set generated by all information available on the market in period t
1_n	$(n \times 1)$ vector of ones
$P_{n,t}$	transaction price of property n in period t
$\text{rank}(A)$	rank of matrix A
s_t	vector with market indicators for period t
$S_{n,t}$	indicated value for property n in period t given by the sales comparison approach
$\tau_{n,t}$	remaining time of usage for property n in period t
$\text{tr}(A)$	trace of a matrix A, i.e., the sum of its diagonal elements
$V_{n,t}$	market value of property n in period t

$V^a_{n,t}$	ascertained market value of property n in period t
$\mathcal{V}[X]$	variance of random variable X
$\mathcal{V}_t[X]$	variance of random variable X conditional on $\mathcal{F}_t$
$\mathrm{vech}(A)$	stags the lower triangular part of the square matrix A into a vector
$x_{n,t}$	row vector of characteristics of property n in period t
0_n	$(n \times 1)$ vector of zeros

List of Figures

2.1 Different stages of the appraisal process 9

3.1 Berlin's districts 72
3.2 Plot of the value of the reference dwelling from 1982:8 to 1999:12 74
3.3 Smoothed expected return deviations. Plot of the estimated series from 1982:8 to 1999:12 80

4.1 Input and output for a price prediction from MD*Immo 97
4.2 Screenshot of MD*Immo's request form 98
4.3 MD*Immo's technical implementation 99

5.1 Average yearly price-rent ratios from 1980 to 2000 114
5.2 Nonparametric density estimates for ratios of price to income value 119
5.3 Nonparametric density estimates for ratios of income value to price 120
5.4 Nonparametric density estimate for log ratios of price to income value 122
5.5 Smoothed general market conditions 1982:6-2000:5 126

6.1 Single-family house prices and construction costs from 1995:1 to 2002:3 137
6.2 Nonparametric density estimates for ratios of price to cost value and for ratios of estimated market value to cost value . 139
6.3 Quarterly average ratios of price to cost value and estimated market value to cost value 140

List of Tables

2.1 Overview on commonly used appraisal approaches in Germany for residential properties . . . 18
2.2 Purposes of asset market indexes . . . 37
2.3 Publicly available indices for real estate in Germany . . . 41

3.1 Summary statistics for transacted single-family houses in the four South-West districts of Berlin between 1982:08 to 1999:12 73
3.2 OLS estimates of optimal specification of the hedonic equation for single-family houses 1982:8 to 1999:12. . . . 76
3.3 Estimates of the parameters in the SSM and influences of the market indicators . . . 78
3.4 Regression results for the process of the rent index from 1980:2 to 1999:12 . . . 88
3.5 Summary statistics for the market indicators and ADF tests for 1982:8 to 1999:12 . . . 90

5.1 Number of observations with appraised income value per year 112
5.2 Summary statistics for transacted apartment houses in Berlin, Germany between 1980:1 to 2000:5 . . . 113
5.3 Summary statistics for ratios of price to income value and for ratios of income value to price . . . 115
5.4 Linear regression for q on object-specific characteristics . . . 116
5.5 Comparison of mean squared prediction errors and mean absolute errors for income valuations . . . 118
5.6 Estimated SSM for the general market conditions for 1982:6-2000:5 . . . 124

6.1 Summary statistics for transacted single-family houses in Berlin between 1995:1 to 2002:3 . . . 133

6.2 OLS estimates of optimal specification of hedonic regression for single-family houses 1995:1 to 2002:3 135
6.3 Summary statistics for ratios of price to cost value and for ratios of estimated market value to cost value 138
6.4 OLS estimates of optimal regression specification for q ratios 1995:1 to 2002:3 141

Chapter 1

Introduction

1.1 The importance of real estate appraisals

> "The role of appraisals cannot be overemphasized because appraised values are used as a basis for lending and investing."
>
> Brueggeman and Fisher (2001, p. 226)

In many countries, real estate is a large component of capital stock and a large component of economic wealth. Real estate is commonly classified into two classes: residential and nonresidential properties (Brueggeman and Fisher; 2001). Residential properties are single-family houses, condominiums, and multi-family properties such as apartments houses. Nonresidential properties are office and retail buildings, factories, warehouses, hotels and institutional real estate like hospitals and universities.

Focusing on residential real estate, figures from statistical agencies give the impression of its importance within capital stock. According to the Federal Statistical Office in Germany, residential real estate structures represented 49% of the capital stock at the beginning of 2001 (Deutsche Bundesbank; 2002). In the United States, according to the Bureau of Economic Analysis, residential real estate structures represented about 39% of the capital stock in 2000 (Herman; 2001). In both cases, the value of real estate structures is the sum of new investment plus the value of the existing stock, where discarded properties are deducted from the stock. The value of the existing stock is calculated at replacement cost for new structures subtracting a net depreciation value for age (Deutsche Bundesbank; 1998; U.S. Bureau of Economic Analysis; 1999). It is important to emphasize that the value of land is not included in these figures. One can derive an estimate of the value of real estate in the portfolios of private households by adding the value of land to the estimated value of structures. For Germany, the share of real estate in the portfolio of households was on average 53% in the nineties (Deutsche Bundesbank; 1999). This shows that residential real estate is the most important asset in the portfolios of many households. Even people that do not own an apartment or single-family house might have shares of real investment funds in their portfolios.

However, estimates of the value of real estate often vary between different studies. It is well-known that the method of valuing the stock of real estate structures might be affected by measurement inaccuracies, because it is difficult to adjust the replacement costs for depreciation or technical obsolescence (Deutsche Bundesbank; 1998). As DiPasquale and Wheaton (1996, p. 4) emphasize, changes in the value of the stock "are a function of market forces that are unlikely to be captured even by the most sophisticated estimates of depreciation." Furthermore, it is difficult to estimate the value of land. Using different valuation approaches, Bach and Bartholmai (1998) derived different figures for the value of residential real estate owned by German households. The varying figures are remarkable, because they reveal that valuation of properties seems to be a difficult task.

Accurate valuations are crucial on the individual level. For households, firms, and government agencies, it is of great interest to know the market value of their property even if they do not have the intention to sell their properties on the market. The market value is the price that their property should bring in the market given usual business dealings. The knowledge of the market value might be necessary for accounting purposes, for decisions based on the composition of wealth or consumption, or for loan applications. Wrong decisions can be made if the assessment of market values is incorrect, because favorable investments are postponed or are done too early.

Whereas it is nearly cost-free to observe announced offer prices for properties in newspapers and on the Internet, transaction prices are difficult to observe. That makes a difference to other assets, e.g., most bonds and stocks are quoted on exchanges so that the market value can be observed instantly. Even if transaction prices of real estate are observable, market values for individual properties are difficult to figure out if the traded properties are only partly comparable or if the observed transactions are outdated.

Given the importance of real estate, it is clear that many approaches to value and predict prices of real estate exist. In Germany, as in many countries, professional bodies of real estate agents provide price indexes that allow access to the average level of real estate prices. Moreover, valuation of individual properties is a profession. In Germany, legislators believe real estate valuation is so important that valuation techniques have been codified in the Regulation on Valuation (WertV) and the Guidelines on Valuation (WertR 91). This is also true for the valuation of real estate that will serve as collateral for mortgage loans. Here, the techniques that are used by banks must be approved by the German Financial Supervisory Authority (BaFin), see the German Mortgage Banks Act (§ 13 HypBankG).

Despite the advantage that German appraisal data are outcomes of a largely standardized process, the empirical knowledge on the accuracy of different approaches to predict individual real estate prices is rather limited. This study enhances the knowledge on German valuation techniques and evaluates their accuracy. It shows that the codified German valuation techniques can be justified by economic models for property markets. Moreover, given this interrelation, appraisal data are a convenient devise for empirical models of real estate markets. Because every valuation technique tries to mimic how prices are established in the market, the study delivers valuable insights into the mechanisms of real estate markets.

1.2 Structure of the study

The codified German valuation techniques are based on valuation approaches used in many other countries: the sales comparison, the income, and the cost approach (Brueggeman and Fisher; 2001, Chap. 8). In this study, we clarify the economic concepts that lie behind the different approaches to predict market values of properties. We evaluate their accuracy with data sets from the real estate market of Berlin that comprise to a large extent the market volume since 1980.

Whereas we have appraisal data for the income approach and the cost approach, we have no data available for the sales comparison approach. But, as we will show, a variant of the sales comparison approach is hedonic regression. We use a regression based approach to model Berlin single-family house prices in a state space model. With respect to the three major valuation approaches, our model is a combination of the sales comparison and the income approach. The online price prediction service MD*Immo is a practical outflow of our state space model. This service can be used to obtain up to date market values of single-family houses in Berlin.

We start in Chapter 2 with an overview on valuation approaches and clarify their concepts. After introducing the major approaches, we give a short overview of the codified valuation techniques in Germany. Moreover, we discuss at length two economic models for residential real estate markets, which serve as references for the whole study. We give an overview of price indices in Germany that should measure the common behavior of real estate prices. There are two major approaches to construct price indices for real estate. The first approach uses appraisals or expert ratings of real estate to construct property indices, and the second approach uses transaction prices to estimate price indices that explicitly control the heterogeneity of the properties. Currently, most of the German property price indices are expert-based. However, expert-based indices can be misleading if they do not control accurately for the heterogeneity of the appraised objects. The regression-based hedonic approach try to cope with that heterogeneity. According to this approach, transaction prices are given by a common market component and weighted characteristics, where the weights are implicit prices for the characteristics. By estimating the common market component and the implicit prices, it is possible to construct a price index for a standard house with fixed characteristics. The chapter presents different hedonic regression models and emphasizes its lack of economic motivation.

In Chapter 3, we study the movement of single-family house prices in Berlin during a span of nearly twenty years. We emphasize that houses are an asset to their owner, where values are given by discounted imputed rents. We derive our hedonic regression equation from the present value, thus providing theoretical motivation and aiding interpretation of the hedonic model and the common price component. This component is associated with expected deviations from the long-run rate of return of single-family homes. We fit the imputed rents with a flexible hedonic function and use the EM algorithm to estimate our state space model of house prices. The empirical results are sensitive and reveal which characteristics and factors influence single-family house prices.

In Chapter 4, we present our online price prediction service MD*Immo for single-family houses in Berlin. MD*Immo is a direct outflow of the state space model for Berlin house prices—presented in Chapter 3—and allows the user to request the market value for a property. For lucidity, the user gives only five characteristics of the subject single-family house. The characteristics are used to calculate the expected price of the house given the estimated implicit prices. Data from 1995 up to the current date are used to estimate the unknown implicit prices. We describe the technique of the request process and the technical implementation of the service in detail.

Chapter 5 deals with income valuation according to the WertV. We clarify the process of income valuation and show that the present value is the economic rationale for the income approach. Short-run deviations between transaction prices and appraisals can be seen as short-run deviations in the discount rates for residential real estate. The short-run deviations are influenced by variables which reflect the current state of the market. We use data on appraisals and transaction prices for apartment buildings in Berlin to evaluate the accuracy of income valuation according to WertV. The outcomes of the income valuation are compared with the outcomes of simple capitalized gross rents. Eventually, we estimate a state space model for the short-run deviations between transaction prices and appraisals. It is common practice that appraisers adjust their final estimate according to a valuation approach for so-called common market conditions. Our state space model allows for the identification of factors that influence the common market conditions. We estimate the magnitude of these influences.

Chapter 6 deals with the cost approach according to the WertV. We show how this approach has to be adopted according to the WertV. The rationale of the cost approach is in accordance with Tobin's Q theory. The theory clarifies that the cost approach has a major drawback: the market can be out of equilibrium for extend periods and cost values will deviate from market values. Whereas the other two major valuation approaches value a property directly, the cost approach values a property only indirectly via the replacement costs. Using transaction prices and appraisals for single-family houses since 1995, we evaluate the accuracy of the cost approach according to WertV. The results suggest that cost values according to WertV are not reliable.

Chapters 3 to 6 are nearly self-containing. Chapter 3 is based on joint work with Axel Werwatz, see also Schulz and Werwatz (2002). The online prediction service MD*Immo described in Chapter 4 is a joint project with Axel Werwatz, Rodrigo Witzel and Hizir Sofyan, where the last two were responsible for the technical implementation.

A final comment is necessary on what the data represent. All data are from Berlin. Before 1990, the data are only for the West part of Berlin. Thus, one can argue that the study is only of limited, regional interest. However, there are two arguments to rebut such an objection. The first argument is that it is quite common to use citywide data to scrutinize the behavior of real estate markets. The assumption is that house prices are formed on local markets and that nationwide influences are small. Poterba (1991) found that citywide data are much more valuable for scrutinizing house price behavior than aggregated data, because the latter aggregate and smooth the data too much. The second argument is that the real estate market of Berlin is exemplary to other German regions. Moreover, the codified German valuation techniques are based on approaches which are used in many countries. Thus, the results of the study are not restricted to a mere regional context.

Chapter 2

Measures of Real Estate Prices

2.1 Valuation approaches

2.1.1 What is a valuation approach?

An appraisal for an individual property is an estimate of its value. There exist different purposes for which appraisals of individual properties are needed: valuation for investment decisions, like buying, selling, building a property, or performance reports; valuation of properties as collateral for lending purposes; valuation of properties for insurance policies and for taxation. The effective date for which the value is ascertained depends on the purpose of the appraisal. Whereas valuation for investment decisions are estimates of the current value at the date of the appraisal, valuations for lending purposes should reflect the value of the collateral when a default by the borrower is likely. In this chapter, we focus on the valuation of the current value of a property for investment decisions. This is no severe restriction because valuations with other purposes—like lending decisions—are often oriented on the approaches discussed in this chapter.

We denote the appraised value at the date of the appraisals as its ascertained market value. According to the U.S. Uniform Standards of Professional Appraisal Practice, the market value is the "most probable price which a property should bring in a competitive and open market under all conditions requisite to a fair sale, the buyer and seller each acting prudently and knowledgeably, and assuming the price is not affected by undue stimulus" (Brueggeman and Fisher; 2001, p. 224). In Germany, the market value (Verkehrswert) is defined in the Building Law as the transaction price that one should expect for an object given its characteristics and usual business dealings (§ 194 BauGB). It is clear that both definitions do not coincide because the most probable price is not necessarily the expected price. Stated in statistical terms and assuming a distribution of possible transaction prices, the Uniform Standards define the market value as the modus of the distribution and the German Building Law defines as market value the expected transaction price.

Despite this difference, both definitions emphasize that all characteristics of a property have to be assessed, and that no unusual circumstances—like a forced sale, or affiliation between seller and buyer—that might occur during the sale should be taken into account. Then, the ascertained market value is the price that one should expect to observe when the property changes hands between a typical buyer and a typical seller in a competitive market. Therewith, typical buyers and sellers value a property only with respects to its characteristics and are not affiliated or forced to bargain with each other. Non-typical sales happen if buyers and sellers are relatives, divorced couples, neighbors or if other unusual circumstances influence the business dealings.

Let $V_{n,t}$ denote the market value of property n in period t, which is determined given the correct market model which explains how market values are established and given all relevant information which is necessary to figure out $V_{n,t}$. This information is collected in the set $\mathcal{F}_{n,t} = \{x_{n,t}, \Omega_t\}$, where the row vector $x_{n,t}$ contains property specific information, like the age, the lot size, and the rent, and Ω_t contains information on the market, like interest rates or the number of building permissions. Unusual circumstances are modelled with the random variable $U_{n,t}$, which is independent from $V_{n,t}$ and has an expectation of one. Let $P_{n,t}$ denote the transaction price of property n in period t, then it follows from the German definition

$$P_{n,t} = V_{n,t} U_{n,t} \ .$$

Multiplicative disturbances $U_{n,t}$ can be justified by the fact that proportional figures are common in real estate business. Moreover, taking the conditional expectation of $P_{n,t}$, one obtains

$$\mathcal{E}_t[P_{n,t}] = V_{n,t} \ .$$

where $\mathcal{E}_t[P_{n,t}]$ is the expected value of $P_{n,t}$ given $\mathcal{F}_t$. All deviations between the market value and the transaction price $P_{n,t}$ occur during the sale due to unusual circumstances.

Appraising by a valuer is a systematic process that results in an estimate of the market value. Figure 2.1 shows the different stages of the appraisal process (Brueggeman and Fisher; 2001, p. 225). In the first stage, the appraiser

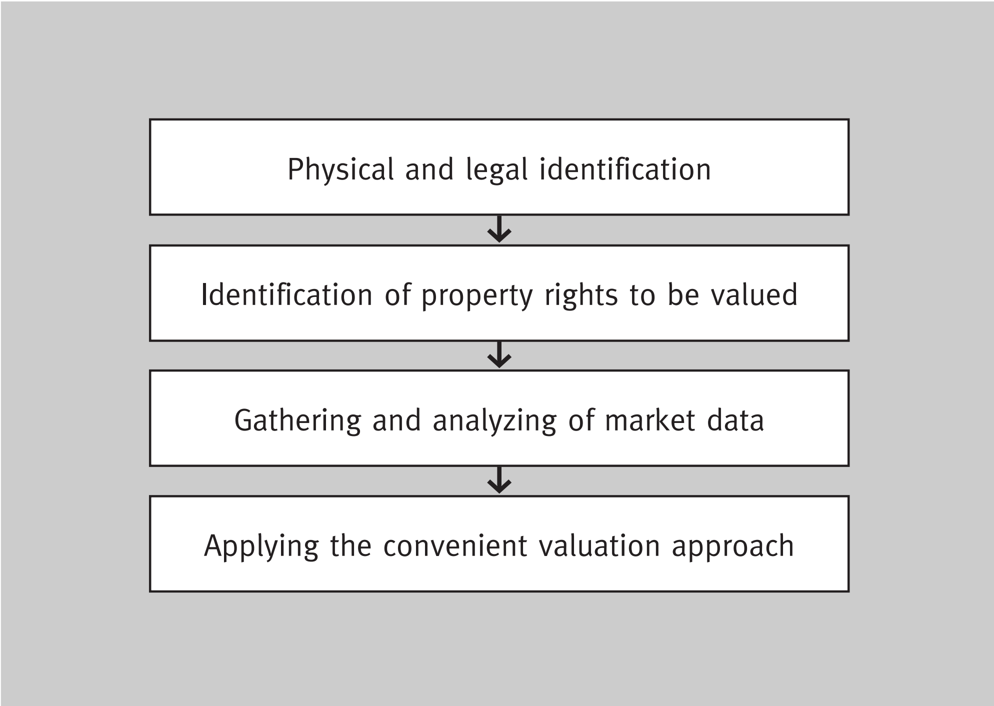

Figure 2.1: Different stages of the appraisal process for estimating the market value.

ascertains the physical structure of the property, i.e., the type of building (for example: apartment house, single-family house) or the type of land (for example: building or farm land), the shape of the lot, the age of the building, and so on. Furthermore, the appraiser has to ascertain legal rights on the property, for example, rights to pass or to use parts of the land by other individuals. In the second stage, the property rights that should be valued have to be identified. The aim of valuation can be the lot with building, the value of the building or a part of the lot. The third step consists in gathering market information like transaction prices of comparable properties, rents, costs, vacancies, interest and mortgage rates. In the last step, the appraiser decides on the approach to estimate the market value given all information the appraiser has gathered on the property and the market. The choice of the valuation approach depends largely on the property type and on the data at hand.

We recognize a valuation approach as a systematic way of providing the value of a property given the information gathered in the first three steps of the appraisal process. We term the output of a valuation for property n as its indicated value. Let $\mathrm{VA}(\cdot)$ be the specification of a valuation approach, then

$$\text{Indicated value} = \mathrm{VA}(\mathcal{F}_{n,t}) \tag{2.1}$$

where $\mathcal{F}_{n,t}$ comprises all information on property n that is being valued, collected in $x_{n,t}$, and all information on the real estate market in period t, collected in Ω_t, which are necessary to carry out the calculations according to the valuation approach. It is assumed that a appraiser can gather the same information as market participants. The appraiser combines the information on the subject property and the market according to the chosen valuation approach to derive the indicated value.

In Germany, the appraiser has to adjust the indicated value to derive the ascertained market value if this is necessary. Let M_t denote the market adjustment factor for period t which reflects general market conditions, then the ascertained market value $V^a_{n,t}$ for the subject property n in period t is

$$V^a_{n,t} = M_t(\Omega_t)\mathrm{VA}(\mathcal{F}_{n,t}) \; . \tag{2.2}$$

It is important to emphasis that the general market conditions are independent of the characteristics $x_{n,t}$ of the subject property. All information on the subject property should be included in the indicated value and adjustment by M_t should reflect only systematic factors.

The major valuation approaches are

- sales comparison approach
- income approach
- cost approach.

We show in the following, that each of the three valuation approaches underlies a market model on the price formation in the real estate market. It prescribes how to combine the information to derive an ascertained market value. We state the economic rationale of the approach and explain which data are necessary for using the valuation approach. However, even after deciding on a valuation approach, there are different techniques for calculating the indicated value of the subject property. Therefore, we present the most important valuation techniques for the different valuation approaches. In order to distinguish the dependence of the indicated value on the valuation approach that is used for its calculation, we make use of the following notation: $S_{n,t}$ denotes the indicated value of the sales comparison approach, $E_{n,t}$ denotes the indicated value of the income capitalization approach, and $C_{n,t}$ denotes the indicated value of the cost approach.

2.1.2 Sales comparison approach

In the sales comparison approach transaction prices of highly comparable and recently sold properties are used to estimate the market value of the subject property being valued. The economic rationale of the sales comparison approach is that no informed investor would pay more for a property than other investors have recently paid for comparable properties given that the general market conditions are the same. If meanwhile the general market conditions have changed, then investors are only willing to pay comparable prices adjusted by the general price level for properties.

For this approach, the appraiser must have several comparable properties on hand. First, the appraiser has to adjust transactions prices of the comparable properties for differences in characteristics with respect to the subject property. For example, some comparable properties might have a higher gross annual rent or a different age. After deriving a set of adjusted transaction prices—individual indicated values—the appraiser has to reconcile the final adjusted transaction price by weighting the individual indicated values.

To clarify the sales comparison approach, we use the model of Isakson (2002). For simple notation, we assume that all transactions of the comparable properties happened in the preceding period $t-1$. Let $P_{j,t-1}$ denote the transaction price of a comparable property with $K > 0$ characteristics that are given by the $(1 \times K)$ vector $x_{j,t-1}$. The appraiser has information on $J > 1$ comparable properties. The transaction prices are stacked in the $(J \times 1)$ vector P_{t-1}, perhaps measured in appropriate units like price per square meter. The respective characteristics are stacked in the $(J \times K)$ matrix X_{t-1}. The K characteristics of the subject property are collected in the $(1 \times K)$ vector $x_{n,t}$. The adjustment factors are collected in the $(K \times 1)$ vector a. Then, the J individual indicated values $S_{n,t,j}$ for the subject property are given by the elements of

$$S_t = P_{t-1} + (1_J x_{n,t} - X_{t-1})a \tag{2.3}$$

where 1_J is a $(J \times 1)$ vector where all elements are 1. Thus, the individual indicated value for the subject property using the transacted property j is just the transaction price of that property adjusted with weighted differences in characteristics.

In the reconciliation step, the appraiser has to weight the J individual indicated values with the $(1 \times J)$ vector of weights w, where the weights sum to one, i.e., $w1_J = 1$, and all weights are positive. When all elements of w are $1/J$, then the individual indicated values are be weighted equally. In that case, the final indicated value is just the mean of the individual indicated values (a weighting scheme often used by German appraisers). The reconciled indicated value is thus

$$S_{n,t} = wS_t$$

and the ascertained market value according to the sales comparison approach is given as

$$V^S_{n,t} = M^S_t S_{n,t} \ ,$$

where M^S_t is a common market adjustment factor that should reflect the current situation on the real estate market. The superscript indicates that the adjustment factor has to correct for conditions that are not already incorporated in the indicated value of the sales comparison approach. In Isakson (2002), $M^S_t = 1$, but we want to reflect the fact that German appraisers have to adjust the indicated value according to the general market conditions in t. According to the general valuation formula (2.1), the Ω_t contains the prices and characteristics of the comparable sales, information on the adjustment factors a and the weights w.

To derive the individual indicated values $S_{n,t,j}$, the appraiser needs adjustment factors a. Often, these factors consists of an educated guess or the expert judgement of the appraiser. There are no unified techniques to derive these factors. This is also true for the weighting factors w that have to be used in the reconciliation step. Generally, the weights w are necessary to derive a final indicated value. Under our simplifying assumption that all transactions happened in the previous period, equal weights $w_j = 1/J$ are plausible. Under more general conditions, where the appraiser has to use transaction observations of earlier periods than $t-1$, the weights w_j may adjust for that fact. Earlier transaction prices can be down-weighted compared to more recent observations.

It is obvious that the appraiser has to rate many unknowns by using the sales comparison approach. In this sense, the traditional technique of the sales comparison approach is not very precise and results will depend on the talent of the appraiser to find reliable adjustment factors.

A variant of the sales comparison approach uses statistical techniques to derive the adjustment factors a; thus, this technique is precise in the method to derive the adjustment factors. Given comparable sales and using multivariate regression, adjustment factors for the value-determining characteristics are estimated and these estimates are used for appraising the subject property. This technique is well-known to economists as the hedonic approach and will be discussed in Section 2.3. Let us show how this technique resembles the sales comparison approach. For convenience, we still assume that all transactions of the comparable properties happened in the preceding period $t-1$ and that the transaction price of property j is a linear function of its characteristics and a white noise term $u_{j,t-1}$. We obtain therefore

$$P_{j,t-1} = x_{j,t-1}\beta + u_{j,t-1} \ .$$

Here, β are unobservable implicit prices for the characteristics. Assuming that the number of observations J is larger than the number of characteristics K and that the rank of X_{t-1} is K, the unobservable implicit prices can be estimated with OLS as

$$b = (X_{t-1}^\top X_{t-1})^{-1} X_{t-1}^\top P_{t-1} \ .$$

The indicated value is given as

$$S_{n,t} = x_{n,t} b \ . \tag{2.4}$$

Assuming that a constant is included in X_{t-1}, so that $1_J^\top (P_{t-1} - X_{t-1}b) = 0$ (Johnston; 1984), we obtain for (2.4)

$$S_{n,t} = \frac{1}{J} 1_J^\top \left\{ P_{t-1} + (1_J x_{n,t} - X_{t-1}) b \right\} \ .$$

This equation closely resembles the one given of the individual indicated values of the standard sales comparison technique (2.3). However, different to the standard technique, there is no need to weight individual indicated values when the regression technique is used. The regression technique assumes that there can exist only a single indicated value, see (2.4). Thus, the regression based sales comparison approach derives the adjustment factors with an clearly stated objective technique that can be used, at least in principle, by everyone. When data of transactions from several periods are used, time dummies can capture general market conditions in the different periods. Once again, this is an objective technique to cope with the history of observed sales. Recall, that the traditional technique is not clear at this point, and appraisers use their own, subjective techniques. Moreover, the estimated time dummies can be used for fitting a time series model for the general market conditions and the model can be used afterwards for predicting the current market conditions and thus the adjustment factor M_t.

2.1.3 Income approach

In the income approach, the forecasted services that flow from the property are discounted to the date of the appraisal. If the property is rented, then the service flow is given by the net rents minus all operating costs. This service flow is called the net operating rent. Properties that generate a flow of net operating rent are called income producing properties. If a property is owner-occupied, then the flow of services is called the imputed rent (less all operating costs). The discounted value gives the indicated value for the subject property. The economic rationale of the income approach for existing properties is that no investor will pay more for a property than he will retrieve by holding the property. This valuation approach is known as the present value or discounted cash flow (DCF) from the investment literature, see for example Brealey and Myers (2000).

Comparable with the sales comparison approach, there exist different techniques for calculating the income value. For the sake of simplicity, we present only techniques that use a constant discount rate R. This rate should reflect the return that an average investor would command for investments with the same risk like an investment in the subject property. In Germany, constant discount rates have to be used for income valuation which motivates our simplification.

Let $E_{n,t}$ denote the indicated value of the subject property according to the income approach and with $D_{n,t+j}$ the net operating rent in period $t+j$, then

$$E_{n,t} = \sum_{j=1}^{\infty} \frac{\mathcal{E}_t[D_{n,t+j}]}{(1+R)^j} \ . \tag{2.5}$$

$\mathcal{E}_t[\cdot]$ denotes the expectation operator given all information up to period t.

According to the general valuation formula (2.1), this information is collected in $\mathcal{F}_{n,t}$. The appraiser uses the information of the property under valuation—for example, the current rent, the current operating cost and the vacancy rate—and relevant information of the real estate market to rate the expected net operating rents for the coming periods. Quite naturally, the current building will obsolesce, so the appraiser has also to rate the net operating rents for a new building that will be built on the site in the future. It is common usage in the practice of valuation to split the right-hand side of the indicated income value (2.5) into two parts

$$E_{n,t} = \sum_{j=1}^{\tau_{n,t}} \frac{\mathcal{E}_t[D_{n,t+j}]}{(1+R)^j} + \frac{W_{n,t+\tau_{n,t}}}{(1+R)^{\tau_{n,t}}} \ .$$

The first expression gives the present value of the discounted net operating rents for the current building which will be used up in $\tau_{n,t}$ periods and the second expression gives the discounted expected resale price that the owner can expect to receive for the site with the obsolesce building on it. The expected resale price $W_{n,t+\tau_{n,t}}$ has to be discounted to the current period.

A variant of the income approach that is in common usage by real estate agents capitalizes the current net operating rent to calculating the indicated value. Assuming a constant expected growth rate G of income, we obtain from (2.5) that

$$E_{n,t} = \frac{D_{n,t}}{\theta} \ ,$$

where

$$\theta \overset{\text{def}}{=} \frac{R-G}{1+G} \ . \tag{2.6}$$

Let $c = 1/\theta$ denote the capitalization factor, the indicated value is calculated by simply multiplying the net operating rent of the subject property with this factor. This valuation approach is also known as the multiple approach and in Germany it is labelled as the real estate agents method (Maklermethode). The capitalization factor is often calculated by averaging the ratio of transaction prices and rents of comparable sales. If only gross rents are observable, then these are used to calculate the capitalization factor and the indicated value of the subject property is given by multiplying its gross rent with this factor.

The last step in appraising the market value of the subjected property consists in adjusting the indicated value $E_{n,t}$ with the general market conditions M_t^E and thus

$$V_{n,t}^E = M_t^E E_{n,t} \ .$$

The superscript E for the general market conditions indicates that the adjustment factor might depend on general influences that are not already incorporated in the indicated value according to the income approach. If investors use time-varying discount rates for discounting net operating rents, then their effect on expected transaction prices has to be incorporated by M_t.

2.1.4 Cost approach

The last major valuation approach is the cost approach. As we have already mentioned in the introductory chapter, this approach is commonly used for the valuation of the housing stock of national economies. Its economic rationale is that no rational investor will pay more for an existing property than it would cost to buy the land and to build a new building on it. However, given that construction of buildings needs time and that land for building purposes might not be immediately available, prices and costs will diverge in the short-run.

Let $L_{n,t}$ denote the value of land with identical characteristics to the land of the subject property and let $B_{n,t}$ denote the costs for building the structure with the same characteristics, then the indicated value according to the cost approach is just given by

$$C_{n,t} = B_{n,t} + L_{n,t} \ .$$

Because the land of the subject property is already developed, the appraiser has to estimate the land value by using the sales comparison approach. The cost of building a structure with identical characteristics is valued by first calculating the replacement costs of a new structure and second by downward adjustments due to physical or functional deterioration. According to the general formulation (2.1) of valuation approaches, in the cost approach $x_{n,t}$ are the characteristics of the property that are necessary to establish the reconstruction costs of the structure and information on the land. Ω_t contains information on building costs, depreciation rates, and information for valuing the land.

After appraising the indicated value $C_{n,t}$, the appraiser has to adjust the indicated value by an adjustment factor for the general market conditions.

Thus, the appraised market value is

$$V_{n,t}^{C} = M_t^C C_{n,t} \ , \tag{2.7}$$

where the superscript C at M_t^C indicates that it adjusts for influences that are not already incorporated in the indicated value according to the cost approach. M_t considers frictions in the availability of land for building purposes and in the construction industry.

2.1.5 Final reconciliation and combinations of the approaches

It is common usage for appraisers to appraise a property with more than one approach. This depends on property type and available information. If more than one approach is used, then the appraiser has to make a final reconciliation step to derive a final estimate of the market value. However, there exists no methodology on how to mix the different market values calculated by different approaches into one final market value. Once again, the reconciliation step depends solely on the expert judgement of the appraiser.

2.1.6 Codified valuation techniques in Germany

All of the above valuation approaches have codified techniques in Germany through the WertV. The accompanying WertR 91 clarify the steps for appraising the market value, explain how different figures—like the net operating rents or the redevelopment costs—have to be calculated and give approximate figures like long-run discount rates and relative shares of average operating costs. The approximate figures have to be used if no information on them is obtainable for the subject property.

The central figure in the WertV is the market value (Verkehrswert). It is—as we have already mentioned—the transaction price that one should expect on average for an object given its characteristics and given the general market conditions (§ 194 BauGB). The determination of market values according to WertV are explicitly prescribed for calculating prices of shares of real estate funds (§ 34 KAGG), for compensations after expropriation (§ 95 BauGB), and as a relative ceiling for forced sales (§ 74a ZVG) (Thomas; 1995; Berens and Hoffjan; 1995). Investor and project developers are free to choose the valuation approach they like for their purposes. However, in many cases they use approaches oriented on the codified approaches from the WertV.

Table 2.1: *Overview on commonly used appraisal approaches in Germany for residential properties.*

Property type	**Sales comparison approach**	**Income approach**	**Cost approach**
Single-family house	for typical, standardized objects, like row and semi-detached houses	only for rented objects where owners have the intention to realize income	✓
Condominium	✓	supportive for rented objects	
Apartment house	supportive	✓	supportive

Notes: A tick indicates the major valuation approach for the respective property type. Sources are Gottschalk (1999, pp. 49–50), § 7 Abs. 2 WertV and 3.1 WertR 91.

Appraising the market value of a property according to the WertV is a two-step procedure. In the first step, one of three allowed valuation approaches has to be used to approximate the indicated value of the property. These are the sales comparison, the income, and the cost approach. The valuer has to give reasons for the approach he uses for appraising a property. The choice of the valuation approach has to be in accordance with the type of property under consideration (§ 7 WertV).

Table 2.1 gives an overview of commonly used approaches for appraising residential properties. Standard approaches are identified by a tick without comments. It is possible, however, to use approaches other than the standard approach. For owner-occupied houses, the cost approach is standard. When there are enough comparable sales, the sales comparison approach can be used. It is clear that the choice of a valuation approach is motivated by the obtainable data. Rents are seldom observed for single-family houses and so the income approach cannot be used. Condominiums are parts of buildings and it might be difficult to figure out the value of a single condominium even if the indicated cost of the whole building can be appraised. That explains why the sales comparison approach is favored for condominiums, especially if they are owner-occupied.

The second step consists in adjusting the indicated value with the general market conditions (§ 7 I WertV). The general market conditions should reflect the current general situation on the market for the subject property (§ 3 III WertV). The situation is influenced by interest rates, the supply of new buildings, and the general situation of the economy (WertR 91 1.5.3).

2.1.7 Which valuation approach should be used?

It is clear that all three valuation approaches are in accordance with economic reasoning. Whereas the income approach considers directly the income-producing potential of a property by discounting expected cash flows, the sales comparison and the cost approach measure the value of a property indirectly by inferring what investors paid for similar recently sold properties and by the costs of replicating the property (Corgel et al.; 2001, p. 297). All three approaches presume that investors are rational and compare the benefits from buying or selling a property with investment alternatives. The alternatives would be a similar asset with an identical risk structure (income approach), the buying of a similar property (sales comparison approach), or building a similar property (cost approach).

A question that comes immediately into mind is, Why are there different approaches? Should it not be the case that one approach is better than the others? A first answer to this question is pragmatic, and depends on the property under consideration. For example, for a singular shopping center, it will be difficult to find comparable sales. On the other hand, it will be relatively easy to obtain reasonable rent figures. In that case, the data restriction favors the use of the income approach. Things are different for owner-occupied single-family houses. In that case, rents are not obtainable and the sales comparison or the cost approach have to be used. This reasoning explains why the WertV prescribes more or less the usage of different valuation approaches for certain types of properties, see Table 2.1. The second answer to this question is, that it is difficult to decide which approach is the best. All three approaches rest upon concepts from economic theory. Thus, they can be justified by models that explain why prices should be related to prices of similar objects, rents, or construction costs.

However, in the end it is an empirical question to scrutinize how the different German valuation techniques perform. Given transaction prices and appraisals, their accuracy can be explored empirically. Moreover, it is of great interest to clarify the last adjustment step by exploring which variables might influence the common market conditions. But, that is also possible with transaction prices and appraisals. With a slight reformulation of the general valuation formula (2.1), we obtain

$$\frac{V_{n,t}}{\mathrm{VA}(\mathcal{F}_{n,t})} = M_t \ .$$

Allowing for idiosyncratic influences, transaction prices can be used as market values. Therefore, given observations on prices and appraisals, the general market conditions can be inferred.

The next section presents models which give economic explanations for systematic short-run deviations between transaction prices and appraisals. These models will be used later on for statistical models that scrutinize observed price to appraisal ratios.

2.2 Models for the real estate market

As we have seen, the three major valuation approaches are all motivated by economic reasoning. One might ask why there are different economic rationales for the same market. Should it not be the case that one rationale is the correct one? This answer neglects the reality that models always simplify the complexity of the world.

As we will show, the three major valuation approaches emphasize different aspects of the real estate market. The income valuation approach focuses on existing income generating properties whereas the cost approach focuses on the supply of new properties. To clarify both concepts, we present a model that is oriented on Summers (1981) and Poterba (1984), which is now the standard model of housing markets in textbooks, see Sheffrin (1996), Romer (1996) and the extended model in Miles (1994, Chap. 2). The model treats housing as an asset in a general equilibrium framework with equalized return rates between assets (Whitehead; 1999, p. 1562). However, the model deals only with homogenous properties and neglects the important aspects of heterogeneity. Although its main application area is the evaluation of taxation and interest deductability on the housing stock, see Bruce and Holtz-Eakin (1999), the model gives valuable insights on real estate markets and deliver explanations for market adjustment factors. However, because of the assumed homogeneity, the results of model are rather limited for motivate the sales comparison approach. So, we present a second model which deals with heterogeneous properties.

2.2.1 Asset market approach

The model assumes that houses are homogeneous and that the total quantity of housing stock is $H(t)$. $R\{H(t)\}$ is the rent for a unit of housing where the rent is decreasing in the stock, i.e. $\partial R(H)/\partial H < 0$. The rent is always characterized as a function of $H(t)$ and should thus not be confused with the required return R. To be aware of this fact and because the model is set up in continuous time, we use r for the required return.

$P(t)$ denotes the price for buying one unit of housing. Households can always decide to buy or to rent a house. If a household decides to buy a house and hold it for one period, the household saves the rental cost $R\{H(t)\}$, but incurs holding costs of

$$(m + \delta + \tau)P(t) ,$$

where m are relative maintenance and repair expenditures, δ is the depreciating rate of the building and τ is the real estate tax rate. At the end of the period, the absolute capital gain from holding a house is given by

$$\dot{P}(t) \stackrel{\text{def}}{=} \frac{\partial P(t)}{\partial t} .$$

It is assumed that the households have perfect foresight and anticipate correctly the capital gains for an investment in a housing unit.

With r as the nominal return rate on an alternative investment, the household will be indifferent between buying or renting a housing unit if the rental costs are equal to the ownership costs, i.e.,

$$R\{H(t)\} = \left\{ r + m + \delta + \tau - \frac{\dot{P}(t)}{P(t)} \right\} P(t) . \tag{2.8}$$

Here, $rP(t)$ are the opportunity costs of investing $P(t)$ in a housing unit and not investing in the asset with return rate r. $\dot{P}(t)/P(t)$ is the relative capital gain of holding the property, which has a negative influence on the holding costs.

An interesting reformulation is given, if we let

$$D\{H(t)\} = R\{H(t)\} - (m + \delta + \tau)P(t)$$

denote the (imputed) net operating income—the net service flow of one housing unit—and by rearranging (2.8) to

$$r = \frac{\dot{P}(t) + D\{H(t)\}}{P(t)} . \tag{2.9}$$

In asset market equilibrium, the return r on the alternative investment must be equal to the return of holding a housing unit. That return is given by the capital gain plus the net operating income divided by the employed capital $P(t)$, where we denote the latter ratio as income yield. Rearranging and using that $D\{H(t)\}$ is implicitly a function of t gives

$$\dot{P}(t) = rP(t) - D(t) \; .$$

Now suppose that the development of $D(t)$ is known and that its average growth is smaller than r, then the solution of the above equation is (Poterba; 1984, p. 733)

$$P(t) = \int_t^\infty D(s)e^{-r(s-t)}ds \; . \tag{2.10}$$

The solution is easily checked by differentiation with respect to t and by using Leibnitz's formula. Equation (2.10) shows that the current price for a housing unit in asset market equilibrium is given by the present discounted value of its future net operating income. However, it is important to mention that $D(t)$ still depends on the stock of housing. Thus, the knowledge of $D(s)$ for $s \geqslant t$ includes the knowledge of $H(s)$, the housing stock.

The supply of new housing units is given by an investment function that is governed by the real price of housing units. The net investment of housing units is

$$\dot{H}(t) = I\{Q(t)\}H(t) - \delta H(t) \; . \tag{2.11}$$

Here, we define the real price $Q(t)$ of housing as the ratio of the nominal price $P(t)$ and the replacement costs of housing $C(t)$, see Bruce and Holtz-Eakin (1999). $I(\cdot)$ is the strictly increasing relative (gross) investment function with $I(1) = \delta$. Thus, if the real price of housing units is equal to the replacement costs, then the investment is equal to the depreciated housing units and the net investment is zero. The investment function implicitly assumes that convex adjustment costs exist for changes in the stock of housing units. Given these costs, deviations between the price for existing housing units and replacement costs induce finite investment, see Romer (1996, Chap. 8). The dependence of the investment function on the ratio of the price of existing housing and their replacement cost is known as Tobin's Q theory of investment. In Chapter 6 we discuss this theory in detail.

Rearrangement of (2.11) gives

$$P(t) = I^{-1}\left\{\frac{\dot{H}(t)}{H(t)} + \delta\right\}C(t) \tag{2.12}$$

with $I^{-1}(\delta) = 1$. Equation (2.12) shows that the price of one housing unit is proportional to its replacement cost, where the factor on proportionality is related to the supply adjustment process.

We obtain furthermore

$$\frac{\dot{P}(t)}{P(t)} = \frac{\dot{Q}(t)}{Q(t)} + \pi \ , \tag{2.13}$$

where π is the inflation rate of the replacement costs. For simplicity, we assume that the inflation rate of replacement costs is constant. The real rent and the real net operating income are denoted as $R^{real}\{H(t)\}$ and $D^{real}\{H(t)\}$, respectively. They are just the nominal figures divided by the replacement cost. Dividing the relation between rents and prices given in equation (2.8) by $C(t)$, plugging in (2.13) and rearranging gives

$$\dot{Q}(t) = (r + m + \delta + \tau - \pi)Q(t) - R^{real}\{H(t)\} \ . \tag{2.14}$$

The housing market is in steady state if

$$\dot{Q}(t) = 0 \quad \text{and} \quad \dot{H}(t) = 0 \ . \tag{2.15}$$

The model has a saddle-point equilibrium, see Sheffrin (1996, pp. 152). In equilibrium, we obtain for the investment equation (2.11) and for the asset market equation (2.14)

$$\begin{aligned} I(Q^*)H^* &= \delta H^* \\ (r + m + \delta + \tau - \pi)Q^* &= R^{real}(H^*) \ , \end{aligned}$$

where it follows immediately from the investment equation that

$$Q^* = 1$$

and from the equation of the asset market equilibrium that

$$R^{real}(H^*) = r + m + \delta + \tau - \pi \ .$$

In steady state, the investment is equal to the depreciation of the housing stock and thus remains constant. The real rent equals the user cost of housing. However, given a positive inflation rate π, which is smaller than r, all nominal figures grow with this rate. Let t^* denote the time point when the steady state is reached and with C^* the replacement cost in t^*, then we obtain with

$$C(t) = C^* e^{\pi(t-t^*)} \quad \text{for} \quad t \geqslant t^*$$

immediately that

$$P(t) = C(t) \quad \text{for} \quad t \geqslant t^* \ . \tag{2.16}$$

Using the arbitrage condition (2.9) which states that in asset market equilibrium the return on the alternative asset must be equal to the capital gain plus the income yield, the capital gains are identical with π and we obtain

$$\begin{aligned} P(t) &= \frac{D(t)}{r-\pi} \\ &= \int_t^\infty D(s) e^{-r(s-t)} ds \qquad \text{for } t \geqslant t^* \ , \end{aligned} \tag{2.17}$$

where

$$D(s) = D(t) e^{\pi(s-t)} \quad \text{for} \quad s \geqslant t \geqslant t^* \ .$$

Results with respect to the major valuation approaches

Our model shows that all three valuation approaches are sensible. Given the fact that all housing units are homogeneous, they have in every instant identical prices. Thus, we obtain as a first result from the model:

RESULT 2.1 (Sales comparison approach). *In the model, the prices for existent housing units $H(t)$ are identical to $P(t)$ in every instant. That is the very idea of the sales comparison approach: identical houses should command identical prices.*

It is clear that our model assumes from the outset that all housing units are identical and command the same price. The following results are of more interest:

RESULT 2.2 (Income approach I). *The asset market equilibrium implies always that*

$$P(t) = \int_t^\infty D(s) e^{-r(s-t)} ds \ ,$$

see (2.10). In every instance, and even if the steady state is not yet reached, the price equals the discounted future stream of income from a housing unit. The supply side of the market—the flow of new investments—is implicitly incorporated in future net operating rents. This resembles the economic rationale of the income approach.

Recall that the capitalization factor $1/\theta$ for the simplified income technique is

$$\frac{1+G}{R-G}\ ,$$

see (2.6). Comparing this with $1/(r-\pi)$ from the steady state reveals that both capitalization rates are equivalent. To see this, just recall that in steady state the net operating rents grows with π and that the G in the numerator comes from the fact that we have used a discrete time model where the first net operating rents accrues from the point when the property was bought. Thus, we obtain a second result for the income approach

RESULT 2.3 (Income approach II). *In steady state, the price of a housing unit is given by*

$$P(t) = \frac{D(t)}{r-\pi} \quad \textit{for} \quad t \geqslant t^*\ ,$$

see (2.17). The price is just the capitalized net operating rents, which resembles the capitalization technique of the income approach.

As we have seen in our overview on the income approach in Section 2.1.3, in practice of valuation, the current net operating rents is often capitalized and the resulting indicated value is adjusted for general market conditions to derive the ascertained market value. Defining

$$M^E(t) \stackrel{\text{def}}{=} (r-\pi)\int_t^{\infty} e^{-\{r(s-t)-g(s)\}}ds \tag{2.18}$$

with

$$g(s) \stackrel{\text{def}}{=} \ln\frac{D(s)}{D(t)}\ ,$$

we can rewrite the present value as

$$P(t) = M^E(t)\frac{D(t)}{r-\pi}\ . \tag{2.19}$$

In steady state, the cumulated growth rate $g(s)$ of the net operating rents is $\pi(s-t)$ and thus $M^E(t) = 1$ for $t \geqslant t^*$. If the steady state is not yet reached, then $M^E(t) \neq 1$ and the capitalized net operating rents is different from $P(t)$. The adjustment factor $M^E(t)$ can be seen as the averaged time varying capitalization rate multiplied with the long-run inverse capitalization rate. The important fact is—due to our assumption of a constant return rate r—that the adjustment factor is a function of the growth rates of the net operating rents. We obtain

RESULT 2.4 (Income approach III). *Let $D^a(t) \stackrel{\text{def}}{=} M^E(t)D(t)$ denote the adjusted net operating rent, where $M^E(t)$ reflects future growth of the net operating rents. Given this interpretation, the price*

$$P(t) = \frac{D^a(t)}{r-\pi}$$

is the capitalized adjusted net operating rent. This gives the rationale of the income valuation technique which prescribes adjustment of the current net operating rent to obtain a long lasting net operating rent and to capitalize the adjusted figure with the long-run capitalization factor.

The important relationship for the rationale of the cost approach is given by (2.12). We obtain

RESULT 2.5 (Cost approach). *In every instance , prices and replacement costs are related due to*

$$P(t) = I^{-1}\left\{\frac{\dot{H}(t)}{H(t)} + \delta\right\} C(t) \ .$$

Let $M^C(t)$ denote the factor in front of $C(t)$, and one sees immediately that the replacement cost are different from the price, except in the steady state. In the steady state, $M^C(t) = 1$ and $P(t) = C(t)$. In all other cases, only adjusted costs $C(t)$ resemble prices. The rationale of the cost approach is sensible, but out of steady sate, the indicated value $C(t)$ must be adjusted with a factor that reflects the general market conditions—here the adjustment process of new investment.

From a conceptional aspect, this result is unfavorable for the cost approach. Whereas the rationales for both the sales comparison and the income approach are fulfilled in every instant, see Results 2.1 and 2.2, the rationale for the cost approach—prices are equal to replacement costs—holds only in steady state. Out of steady state, adjustment of the subject value from the cost approach is a necessary prerequisite. Nevertheless, given the whole information on the future evolution of the stock and the rents, the price of a housing unit is totally determined by the investment equation. In conclusion, prices and costs are connected by a functional relationship, but prices are not equal to replacement costs in every instance.

Limitations of the model

The above presented model delivers valuable insights into the rationales of the different valuation approaches, but it has certain limitations. The most obvious one is that houses are treated as homogeneous. Furthermore, two important aspects with respect to land are not treated. The first aspect is the supply of land, which will depend on the relative price of its usage for housing and alternative purposes, and on its total available amount that might be fixed. Poterba (1984) sheds some light on this with the Cobb-Douglas production function for housing services. With this extension, rents comprise remuneration for the usage of building and land and the qualitative results remain the same. The second aspect that has not been treated up to now is the location value of land. For example, houses with a good infrastructure or with low commuting costs will command higher prices per unit compared with houses that are far away from employment centers.

The next paragraph presents a model that explains the location value of land. Given different locations, houses are no longer homogeneous and prices will be different for houses at different locations. However, as we will show, the rationales of the major valuation approaches remain valid. Most attention will be given to the rationale of the sales comparison approach.

2.2.2 Location approach

To illustrate the location value of land, we refer to results from a monocentric circular city model, see DiPasquale and Wheaton (1996, Section 2) and O'Sullivan (2000, Chapter 8).

In the city live n households with the same income and every household demands housing on one unit of land. The area of the city is $b^2\pi$ units of land where b is the radius of the circular city. D^a is the rent for land if it is used for agricultural purposes and it is assumed that this opportunity value is constant throughout the city. Thus, land usage for housing purposes generates opportunity costs. c are the per period replacement costs for the homogenous buildings (structures), where c is independent of the location of the building. In the city, employment opportunities are given only in the center. Every point in the city is characterized by its location $l \in [0, b]$, which is the distance from the city center. A household that lives $l > 0$ from the center incurs commuting costs kl with $k > 0$.

Rents in spatial equilibrium are given by

$$D(l) = D^a + c + k(b^* - l) \ , \tag{2.20}$$

where $b^* = \sqrt{\pi/n}$. Any operating costs are ignored and so rents and net operating rents are equal. In spatial equilibrium, the area of the city comprises $\pi\left(b^*\right)^2 = n$ lots with one unit of land. Thus, every household has a lot to live on. $D(l)$ is the spatial equilibrium rent function because no household has an incentive to change location. Relocation by a distance of Δl changes the commuting costs by $k\Delta l$, and the change in the rent is $-k\Delta l$, and thus the total effect is zero. Equivalently, the costs of renting and commuting $D(l) + kl$ are identical to $D^a + c + kb^*$ in every location. Furthermore, at the edge of the city, the rent minus the per period replacement costs is equal to the opportunity costs D^a of land, so no land owner at the edge has an incentive to change the current usage of land. This is also true for all land owners that have a piece of land inside the city. They make a profit of $k(b^* - l)$ from using the land for housing purposes.

Prices $P(l)$ of houses as a function of the location l are given by the discounted rents

$$P(l) = \frac{D(l)}{R} \ ,$$

where $R > 0$ denotes the discount rate. Let us define the total building costs B and the value of land $L(l)$

$$B \stackrel{\text{def}}{=} \frac{c}{R} \quad \text{and} \quad L(l) \stackrel{\text{def}}{=} \frac{D^a + k(b - l)}{R} \ ,$$

then we obtain

$$P(l) = B + L(l) \ .$$

In equilibrium, prices are the discounted rents and the building plus land costs. Thus, once again, the rationale of the income and the cost approach are resembled. However, prices are now an explicit function of the location, and prices of two houses are only identical if their location is identical. It is easy to derive an explicit relationship between prices of houses with different location l_1 and l_2

$$P(l_1) = P(l_2) + (l_1 - l_2)\left(-\frac{k}{R}\right) .$$

This equation reflects the formula (2.3) for the indicated value of the sales comparison approach. The price for a house at location l_1 is the price for a house at location l_2 adjusted for the evaluated difference in the location.

A further restrictive assumption of all models presented so far is that houses are homogeneous and have the same size of the lot and the same structure. It is easy to extend the analysis to the case where households will choose different sizes of lots and structures. We still assume that the ratio of building size and lot size is fixed. Thus, we assume that the plot ratio—i.e., the ratio between floor space and lot size—is fixed for the whole city due to zoning ordinance, and that it is optimal to exploit it fully.

Let $U(s, x)$ denote a quasi-concave utility function, where s is the size of the lot, which is proportional to the size of the building, and x is a composite of other consumption goods. The household has income y and thus

$$x = y - D(l)s - kl .$$

Here, the rent per unit of land with corresponding unit of building

$$D(l) = c + D_L(l) \tag{2.21}$$

is split into the rent c for the structure and the rent for land $D_L(l)$. Given the location l, the household chooses s to maximize its utility. The optimal size $s(l)$ is given implicitly by

$$\frac{\frac{\partial U}{\partial s}}{\frac{\partial U}{\partial x}} = D(l) . \tag{2.22}$$

The spatial equilibrium is characterized by three conditions

$$\begin{aligned} U\left\{s(l), y - D(l)s(l) - kl\right\} &= \overline{u} \quad \text{for all } l \in [0, b^*] \\ \int_0^{b^*} \frac{2\pi l}{s(l)} dl &= n \\ D_L(b^*) &= D^a . \end{aligned}$$

The first condition guarantees that household utility is the same in every location. The second condition guarantees that all households have a lot. $2\pi l dl$ gives the area of a small ring with distance l from the city center. $1/s(l)$ gives the number of households that have their lot on that ring. Integrating over the whole radius gives the total number of households that must be equal to n. The third condition guarantees that the opportunity costs of land usage for housing is equal to the agricultural yield.

It is easy to see from the first condition and by using (2.22) and (2.21) that

$$\frac{\partial D_L}{\partial l} s(l) = -k \; . \tag{2.23}$$

This characterizes the spatial equilibrium, because no household has an incentive to change its location. If the household moves towards the center by $\Delta l < 0$, it saves $k\Delta l$ commuting costs but has to pay a higher per unit land rent of equal amount. The total decrease in the land rent compensates the household for the decrease in its income and thus holds the real income of the household constant.

The linear spatial equilibrium rent function (2.20) is just a special case of the above given general model. To show this, we consider that $s = 1$ is independent of l. In that case, the differential equation of the spatial equilibrium condition (2.23) has the solution $D_L(l) = A - kl$. It follows from $D_L(b^*) = D^a$ that $A = D^a + kb^*$ and thus $D(l) = c + D^a + k(b^* - l)$.

The total rent that a household pays who lives in distance l from the center is

$$D^T(l) = D(l)s(l) \; .$$

One obtains that

$$\frac{\partial D^T(l)}{\partial l} = D(l)\frac{\partial s(l)}{\partial l} - k \; . \tag{2.24}$$

The total rent function is not necessarily falling in l. The total effect depends on the magnitude of $\partial s/\partial l$ times the rent. $\partial s/\partial l$ is the substitution effect—recall that the land rent function compensates for changes in location—and is positive because the rent per unit of land decreases with l. So the size of lots and buildings increase the farther they are removed from the center.

Once again, prices are given by discounting the rents

$$\begin{aligned} P(l) &= \frac{D^T(l)}{R} \\ &= \frac{\{c + D_L(l)\}\, s(l)}{R} && (2.25) \\ &= B(l) + L(l)\ , && (2.26) \end{aligned}$$

where the building costs $B(l)$ are the discounted per period cost $cs(l)$ and the land costs are the discounted pure land rent $D_L(l)s(l)$.

To give a simple concrete example, let us assume that households have a homothetic utility function. In that case

$$s(l) = \frac{\alpha(y - kl)}{D(l)}\ ,$$

where α is the exponential weight of s in the utility function. The corresponding weight for x is $(1 - \alpha)$. We can decompose the change in s given an increase in the distance l

$$ds(l) = \left(\frac{\partial s}{\partial D} + \frac{\partial s}{\partial y} s\right) \frac{\partial D(l)}{\partial l} dl\ ,$$

where the first expression on the right hand side gives the substitution effect of a rent increase according to the Slutzky decomposition. It is easy to check that the above decomposition holds for our example with

$$ds(l) = (1 - \alpha)\frac{k}{D(l)} dl\ .$$

We obtain for the change of the total rent

$$\frac{\partial D^T(l)}{\partial l} = -\alpha k\ .$$

So, the size of lots increases for location away from the center, but the total rent decreases. The price of a house at location l is just

$$P(l) = \frac{\alpha(y - kl)}{R}\ .$$

Let us summarize the results. In the extended monocentric city model, every house—i.e., lot and building—has two characteristics: its lot size and its location. Using the model, it is possible to show that the formula (2.3) of the indicated value from the sales comparison approach is reasonable. Just take the logarithm of the middle line in (2.25) and assume once again two houses that are $l_2 - l_1 \neq 0$ away. We obtain

$$\ln P(l_2) = \ln P(l_1) + \{\ln D(l_2) - \ln D(l_1)\} + \{\ln s(l_2) - \ln s(l_1)\} \ ,$$

where the transformed price for the house under valuation is just the transformed price of the house with observed transaction and the difference in location—here, the difference in rent per unit land—and the difference in size. The adjustment factors are both equal to one. However, due to the fact that both rent and size are functions of the location, its also possible to state a price relationship with respect to this variable. A first-order Taylor approximation gives

$$P(l_2) \approx P(l_1) + \frac{\partial P(l_1)}{\partial l_1}(l_2 - l_1) \ .$$

For the price function from our example, the adjustment factor is fixed and

$$P(l_2) = P(l_1) - \frac{\alpha k}{R}(l_2 - l_1) \ .$$

The adjustment factor is quite similar to the one from the linear model—which was just $-k/R$—but now the coefficient α from the households utility function is present.

2.2.3 Summary on the rationales of valuation approaches

The section has shown that models of the real estate market support the rationales of the three major valuation approaches. The dynamic model of Summers (1981) and Poterba (1984) has shown that house prices in every case are equal to discounted net operating rents. We have presented reformulations which split discounted net operating rents into discounted current net operating rents times an adjustment factor, which resemble common techniques according to the income approach. Prices in steady state are also equal to replacement costs, but out of steady state, costs and prices deviate. This resembles the technique of the cost approach, where adjustments of cost values might be necessary. Furthermore, we have shown with a simple model of a monocentric city that prices depend on location and that differences in prices due to location can be expressed by adjustment factors times the location difference. This resembles the very idea of the sales comparison approach.

Still, our models have neglected aspects that are important for valuation of properties. The first aspect is that buildings are seldom homogenous, and differences are not only due to size and location. Houses differ furthermore in age, in the number of storeys, in style, and conditions. Thus, the price of a house is not only affected by general market conditions—like the discount factor or the opportunity costs of agricultural land—but also by the characteristics of the respective house. These characteristics have to be considered *and* evaluated in the appraisal process. The second neglected aspect is uncertainty and time-varying discount rates. Despite these inherent simplifications, the models deliver valuable insights into price formation at property markets and underline that the three major valuation approaches are sensitive. Moreover, they reveal that house prices depend not only on property characteristics like the location and the size, but are also driven by common factors like returns for alternative investments and opportunity costs of land usage.

2.3 Index numbers for real estate

The economic models presented in Section 2.2 of the previous chapter have shown that prices of relatively homogeneous property types may be driven by common economic factors. Therefore, instead of concentrating only on the market values of individual properties, it is also interesting to figure out the general behavior of real estate prices. Filtering out the general tendencies of real estate prices is exactly what real estate price indices try to do. Thus, whereas valuation approaches concentrate on the appraisal of individual properties, market price indices measure the common behavior of real estate as an asset class.

This section starts with a short overview on consumer price indices. Then we present purposes of market indices, where we concentrate on real estate price indices. Afterwards, we discuss measurement problems with respect to real estate price indices, which are closely related to measurement problems of consumer price indices.

2.3.1 Market indices for assets

The use of market indexes follows from the tradition of consumer price indexes. Consumer indexes are in use for economic welfare comparisons and as indicators for monetary policy. However, they can also serve as instruments for contract settlements. For example, in Germany it is possible to use the official price index as an escalator in lease contracts, see the German Civil Code (§ 557b BGB).

The most prominent consumer price indices are the Laspeyres index and the Paasche index (Gravelle and Rees; 1992). Let $P_{j,t}$ denote the price for product j in period t and with $Z_{j,t}$ the quantity of that product, then the Laspeyres price index is given as

$$\mathrm{LP}_{0t} = \sum_{j=1}^{J} \frac{P_{j,t}}{P_{j,0}} w_{j,0} \ , \tag{2.27}$$

where

$$w_{j,t} \stackrel{\text{def}}{=} \frac{P_{j,t} Z_{j,t}}{\sum_{k=1}^{J} P_{k,t} Z_{k,t}}$$

is the budget share of good j in period t and J is the number of included goods. The Paasche price index is

$$\mathrm{PP}_{0t} = \left(\sum_{j=1}^{J} \frac{P_{j,0}}{P_{j,t}} w_{j,t} \right)^{-1} \ .$$

Another prominent price index is the Fisher (1922) ideal price index

$$\mathrm{FI}_{0t} = \left(\mathrm{LP}_{0t} \mathrm{PP}_{0t} \right)^{-0.5} \ .$$

The formula of the Laspeyres index is just the ratio of the costs of the bundle $\{Z_{j,0}\}_{j=1}^{J}$ in the current period t to the costs of the same bundle in the base period. Thus, the nominator gives the amount of money that is necessary in period t to buy the same bundle as in period 0. However, due to the fact that consumers will substitute the quantities of goods, the Laspeyres index —which neglects substitution between goods—tends to overstate the change in the cost of living (Diewert; 1998). An important strand of index number theory copes with defining true cost of living index numbers. A true cost of living index relates the minimum cost for archiving the same utility as in the base period, given current prices, to the cost in the base period (Konüs; 1939). To calculate such index numbers, it is necessary to know the preferences, the utility function, of the respective household. As Diewert (1976) has shown, the Fisher ideal index is exact, i.e., a true cost of living index, if the respective household has a quadratic utility function. Furthermore, the quadratic utility function underlying the Fisher ideal can approximate other homothetic utility functions and is thus, using Diewert's terminology, superlative. One can find this strand of index number theory as the economic approach because it evaluates the economic content of index numbers, especially the question of whether or not certain index numbers can be justified by underlying utility functions.

An economic approach is also of great importance to indices that measure the evaluation of asset prices. Such indices play an important role in finance and should have a meaningful interpretation as proxies for market portfolios. For example, such an interpretation exists for the Laspeyres index: assume, that the prices are from assets that an investor holds in quantities $\{Z_{j,0}\}_{j=1}^{J}$ between the periods 0 to t and assume further that these assets pay no dividends. Let

$$1 + R_{j,0t} = \frac{P_{j,t}}{P_{j,0}} \ ,$$

then we obtain

$$\mathrm{LP}_{0t} = \sum_{j=1}^{J} (1 + R_{j,0t}) w_{j,0}$$

and the Laspeyres index gives the total return of the portfolio, where the weight $w_{j,0}$ is the budget share in the portfolio of the asset j in the base period.

Many prominent stock indices are oriented on the Laspeyres price index, such as the Standard and Poor's 500, the Nasdaq, the London FTSE (Gourieroux and Jasiak; 2001, Chap. 15.2.3), and the German DAX (Deutsche Börse; 2002). They weight the included stocks proportional to its market capitalization. The failure of the first index futures market might be an example for the importance of economically sensible asset price indices. In 1982, the market was created at the Kansas City Board of Trade and the traded index was a geometric index of stock prices (Dubofsky; 1992, pp. 248). Although there are competing explanations for the failure—for example, inexpertness with index contracts—, one explanation is based on the fact that a geometric average of numbers not all the same is less than the arithmetic average. Thus, the index could not be physically replicated and was irrelevant for tracking real portfolios (Shiller; 1993a, p. 120). Different to that, all of the above mentioned prominent stock indices are market value-weighted indices, where the underlying portfolio can be replicated. A useful classification of purposes of market indexes is given in Table 2.2, which follows Gourieroux and Jasiak (2001, Chap. 15.2). The classification can be directly translated to real estate. Real estate price indexes should convey information on changes in the value real estate because it is important for investment strategies, important to the potential buyers and sellers of real estate. Such an index should reveal the current state of the market and should measure the evolution of real estate prices. Such indexes are also important for banks that make home loans or use real estate as collateral (Purpose I).

A real estate price index is also of great importance to investors that participate in the real estate market via investment trusts. Such trusts allow investors that can not afford their own real estate property to take part in the real estate market. Investors want to compare the performance of their shares in a trust with the general performance of the real estate market (Purpose II).

Given the fact that real estate comprises the largest component of private wealth, there are some proposals on how to use real estate price indexes to hedge the risk of this portfolio component. Whereas large companies can hedge their real estate risk by portfolio diversification, this is mostly impossible for the average house owner. Given that the value of a house is the major component of the most household's wealth, there is a strong interest in hedging downward slumps in value. Shiller (1993b) proposes futures contracts based on real estate price indexes as a means of hedging against unfavorable price changes. Insurance companies could offer contracts that provide protection against adverse price changes in the same way as they offer fire policies. Insurance companies could hedge their aggregate risk from such value-protecting policies on futures or option markets for real estate prices indices (Shiller; 1993a; Case et al.; 1993). In a recent study in Sweden, Englund et al. (2002) found that the value of hedging possibilities is large, especially to poorer homeowners exposed to house price risk. However, up to now, no index-based contracts are traded that would allow for hedging real estate price risk (Purpose III).

Accurate measurement is also a necessary condition for economic models that try to explain the behavior of households and investors. Researchers are trying to explain the behavior of informed investors and households, and in order to do this, it is necessary that they can use reliable information on real

Table 2.2: *Purposes of asset market indexes.*

I Measures of asset price evolution

Objectives: Indicators for sign and size of price modifications.

Requirements: It has to be computationally simple and evaluated in practice from a limited sample of assets that quickly respond to shocks (i.e., are highly liquid).

II Benchmarks for portfolio management

Objectives: Given the interpretation that a market index conveys the value of an efficient portfolio, the market index is seen as a benchmark for actively managed fonds.

Requirements: The chosen indexes have to be representative for the market and has to include the most relevant assets. Furthermore, it has to measure not only prices changes, but has to include all dividends that occur during the holding period. Indexes of this type are called performance indexes. For example, the German stock index DAX is of this type.

III Support of derivatives

Objectives: Options and futures on market indexes are used as hedging instruments against market risks and for speculative purposes.

Requirements: The chosen indexes have to be updated very frequently and have to be representative for the market to cover the volatility and the composition of the market.

IV Economic indicator

Objectives: Market indexes summarize the value of the underlying assets and can be used for testing and fitting models in the fields of economics and finance.

Requirements: The chosen indexes have to be useful for the models under consideration and have to convey the information that is needed for fitting and testing these models.

Note: Classification according to Gourieroux and Jasiak (2001, Chap. 15.2).

estate prices (Purpose IV).

2.3.2 Difficulties in calculating real estate indices

Different to actively traded bonds and shares, real estate prices are observed mainly for different and heterogeneous properties. Thus, whereas one can observe prices of publicly listed shares on every trading day, it is rare to observe consecutive transactions of identical properties. Therefore, real estate price indexes have to cope with the heterogeneity of transacted properties and with the fact that transacted properties may not be representative of the whole market. Still, problems with heterogeneous products and representative price quotes pertain also to other index numbers. To see this, we present the Laspeyres consumer price index as example. After that, we show that measurement problems for real estate price indices are quite similar.

Regarding the Laspeyres consumer price index

$$\mathrm{LP}_{0t} = \sum_{j=1}^{J} \frac{P_{j,t}}{P_{j,0}} w_{j,0} \ ,$$

the weights $w_{j,0}$ are calculated with data from household surveys and are fixed for several periods, for Germany see Deutsche Bundesbank (1998). Thus, for calculating the Laspeyres price index for period t, only the relative prices $P_{j,t}/P_{j,0}$ are necessary. It is common practice to use more than one price observation to calculate the price for a product because the same product will have different prices in different outlets (Deutsche Bundesbank; 1998; Diewert; 1998). Otherwise the index is not representative.

For the aggregation of price quotes from different outlets, elementary price indices are used. An elementary price index measures the "average" behavior of prices for a given product over time. Let N denote the number of price quotes for a product and let $P_{n,t}$ denote the price for the product charged by retailer n in period t. Then examples of elementary price indexes are the Carli index

$$I_{0t}^{C} = \frac{1}{N} \sum_{n=1}^{N} \frac{P_{n,t}}{P_{n,0}} \ , \tag{2.28}$$

that is the arithmetic mean of the price ratios, the Jevons index

$$I_{0t}^{J} = \prod_{n}^{N} \left(\frac{P_{n,t}}{P_{n,0}} \right)^{1/N} \tag{2.29}$$

that is the geometric mean of price ratios and the Dutot index

$$I_{0t}^{D} = \frac{\frac{1}{N}\sum_{n=1}^{N} P_{n,t}}{\frac{1}{N}\sum_{n=1}^{N} P_{n,0}} \tag{2.30}$$

that is the ratio of average prices. There are several problems that can occur by calculating elementary price indexes

- due to changes in the structure of retailing, the product will be cheaper in new discount stores where prices are not sampled (**outlet substitution bias**)
- prices for a product might not be reported in some periods (**missing observations**)
- products might be driven from the market or are offered only with different characteristics (**quality change and new goods bias**).

Given that the prices are always collected in the same stores, a general shift from high to low price retailers will not be picked up by the index. Thus, the price collection should be in accordance with the whole population of retail stores. The second problem occurs when respondents do not report constantly or if products are sold in the marketplace only in certain periods. The last problem is quite similar to the problem of missing observations and the substitution bias. Either old products are out of the market and prices can not be observed, or products run out of fashion because new products are used instead.

Translated to real estate prices, index construction is exposed to all three problems. Given that only a sub-sample of transacted properties is used for index construction, the index may not be representative of the market. Even if all transacted properties are used for index construction, the index may not be representative of the whole stock, which also includes all non-transacted properties. The problem of missing observations can occur in periods of "cold" markets, where no properties are transacted and no prices are observed. This problem can also occur if the index is constructed for specific properties—like properties held in the portfolio of a large investor—where the properties are not sold in the market. The last problem is the severest one, because it deals directly with the heterogeneity of transacted properties. It would be ideal to have matched observations for every period of index construction because then the index would be calculated for identical properties. In this case, one of the above presented elementary price indices could be used for calculating the market index. But transacted properties will be different from period to period, so per-period matching is impossible. Whereas quality change might be an exceptional problem for consumer price indices, it is the rule for any real estate price index.

Given that real estate consists of heterogeneous properties, there are three ways to produce real estate price indexes

- **expert based indexes**, where well-informed market participants convey their opinion about the prices, perhaps by using transaction data that are known to them because they are often incorporated in the business dealings
- **data-driven statistical indexes**, where transaction data and statistical methods—combined with economic models—are used to estimate the prices of real estate
- **substitute indexes**, for example stock indexes of companies which have their main business in real estate.

2.3.3 Price indices for real estate in Germany

In Germany, the most prominent real estate price indices are reported by professional bodies of real estate agents (Ring Deutscher Makler, Verband Deutscher Makler) and private research firms (Bulwien AG, GEWOS, DID). Table 2.3 gives an overview.

Table 2.3: *Publicly available indices for real estate in Germany.*

RDM Preisspiegel

Objects: single-family houses (detached and row houses) and condominiums in different qualities (simple, medium, good, for detached houses also very good), commercial real estate with vintages before and after 1948.

Type: figures are provided for major German cities and give average prices for single-family houses, average prices per square metre for condominiums, capitalization factors—i.e. price divided by net rent—for commercial real estate.

Frequency: yearly figures.

Data: transaction prices reported by real estate agents for the first quarter of a year. Prices are aggregated by expert knowledge.

Provided by: Ring Deutscher Makler RDM, Federal Organisation of Estate Agents and Property Management e.V.

VDM Immobilienpreisspiegel

Objects: single-family houses (detached, semi-detached and row houses) and condominiums in different qualities (simple, medium, good, very good), commercial real estate.

Type: figures are provided for major German cities and give ranges of prices for single-family houses, ranges of prices per square metre for condominiums, capitalization factors—i.e. price divided by net rent—for commercial real estate.

Frequency: yearly figures.

Data: price ranges reported by members of the VDM, where members use their knowledge of the market and transaction prices of sales where they were involved. Ranges are aggregated by expert knowledge.

Provided by: Verband Deutscher Makler VDM, Organization for Property Management and Financing e.V.

Bulwien Immobilienindex

Objects: residential and commercial real estate.

Type: figures are provided for Western and Eastern parts of Germany. Indices are calculated by weighting city-wide average prices for different types of residential and commercial real estate by the number of inhabitants of the respective city. Indices are calculated by deflating the current figure by the figure of 1975 (West) and 1992 (East).

Frequency: yearly figures.

—continued on the next page—

Table 2.3 *continued*

Data: prices reported by real estate agents, project developers, banks and surveyor commissions for 50 cities in former West Germany and for 10 cities in former East Germany.

Provided by: Bulwien AG.

DEIX Deutscher Eigentums-Immobilien-Index

Objects: single-family houses and condominiums.

Type: figures are provided for Western and Eastern parts of Germany and give average prices per year, deflated with average prices of the year 1995.

Frequency: yearly figures.

Data: transaction prices collected by local surveyor commissions and prices assessed by GEWOS, a consulting institute.

Provided by: ifs Städtebauinstitut in cooperation with GEWOS.

DIX Deutscher Immobilien Index

Objects: commercial real estate that is held in portfolios of professional investors. Index represents in 2001 about 30% of the institutional market.

Type: performance index for institutional real estate market. Different sorts of property are weighted by their total value.

Frequency: yearly figures.

Data: appraised market values of properties held in institutional portfolios. Market values are reported by investors that contribute to the DIX. The figures are checked for validity and plausibility.

Provided by: DID Deutsche Immobilien Datenbank GmbH.

Notes: Sources are Ring Deutscher Makler (1989-1999), Verband Deutscher Makler (2001), Bulwien (2001), Institut für Städtebau, Wohnungswirtschaft und Bausparwesen (2001), DID (2002). Additional information collected from home pages of the respective organizations and by personal communication.

Whereas the first four providers calculate indices that should represent "average" prices for real estate, the DIX is a family of performance indices for the portfolio of contributing institutional investors, where indices are calculated for different property types. The calculation method is the same for all property types. Due to the fact that the properties are held by institutional investors, the DIX is calculated with the ascertained market values of these properties, see DID (2002). The appraised market values are provided by the institutional investors, where the appraisals are calculated according to WertV or the Red Book Definition of the Royal Institution of Chartered Surveyors (RICS). The total return for property n that is held during the years $t-1$ to t is defined as

$$R_{n,t} = \frac{V_{n,t} - (V_{n,t-1} + I_{n,t} - D_{n,t})}{V_{n,t-1} + 0.5 I_{n,t} - 0.5 D_{n,t}} \ . \tag{2.31}$$

Here, $I_{n,t}$ are net investments or dis-investments in property n from year $t-1$ to t. Thus, the numerator gives the capital appreciation of the property plus the net operating rent that accrues to the owner during the year minus net investment expenditure. The denominator is given by the average employed capital during the year.

The DIX total return formula for all N_t properties is

$$DIX_t = \sum_{n=1}^{N_t} s_{n,t} R^a_{n,t} \ , \tag{2.32}$$

where

$$s_{n,t} \stackrel{\text{def}}{=} \frac{V^a_{n,t-1} + 0.5 I_{n,t} - 0.5 D_{n,t}}{\sum_{i=1}^{N_t} (V^a_{i,t-1} + 0.5 I_{i,t} - 0.5 D_{i,t})}$$

is the money-weight of property n and $R^a_{n,t}$ is calculated with ascertained market values. The money-weights are measured with the average employed capital during the year. It is easy to see that

$$DIX_t = \frac{\sum_{n=1}^{N_t} \{V^a_{n,t} - (V^a_{n,t-1} + I_{n,t} - D_{n,t})\}}{\sum_{i=1}^{N_t} (V^a_{i,t-1} + 0.5 I_{i,t} - 0.5 D_{i,t})} \tag{2.33}$$

Thus, the total return of the DIX is given for a notional portfolio that comprises all appraised properties. The DIX total return gives the performance of this portfolio for one year, where portfolio management and valuation fees are neglected. We should mention that the DIX is in effect a Laspeyres price index, see equation (2.27), where the returns $R^a_{n,t}$ are weighted with the average share of property n in the portfolio during the holding period, which is one year.

It is important to stress that the reliability of the DIX as a performance measure crucially depends on the accuracy of the ascertained market values. Whereas the investments $I_{n,t}$ and the current net operating rent $D_{n,t}$ are observable, the market values are not. A severe reliability problem exists if the chosen valuation approach to ascertain market values is inaccurate. The DID controls for the accuracy of the appraisals by requesting the raw data used for the appraisals—like the discount rate or the net operating rent—and demands further justification from the respective appraiser if he uses raw data outside the 90% range for comparable properties. Although this is a consistency check of different appraisers, it does not check the consistency of the used valuation approach. If the valuation approach uses the wrong or an over-simplified market model, then even the best raw data will deliver wrong ascertained market values. As we have already mentioned, many appraisals for the DIX are conducted with the income approach according to WertV. Thus, the empirical study in Chapter 5 is of great interest for index construction.

Whereas the idea of the DIX is evident and the appraisals use common valuation approaches, the calculation principles of the other mentioned German real estate price indices are less obvious. All of them are more or less expert based indices and can be subsumed under the sales comparison approach. However, one has to be cautious by interpreting the term *average* in the above given Table 2.3 as the explicit arithmetic mean of transaction prices. In most cases, the figures are obtained by implicit methods and discussion by expert working parties. The private research firms are especially silent about the methods they use for calculating their price indices.

2.3.4 Summary on German real estate indices

We have seen that the calculation of real estate price indices is hampered by the fact that identical properties are rarely transacted. Appraisal based indices try to circumvent the lack of observed transaction prices for the properties that should be included in the index by using appraisals for these properties. When the sales comparison approach is used for valuation, the lack of comparable transacted properties is still a problem. However, the income and the cost approach do not need information on comparable sales and are thus favorable for appraisal based indices if no comparable sales are at hand. Comparable sales are also important for expert-based price indices, where well-informed market participants—appraiser and real estate agents—convey their opinion about the prices. It is important to emphasize that neither of the above presented German real estate indices is calculated with hedonic regression, that controls explicitly for characteristics of the transacted properties. The next section presents hedonic regression, which is the most prominent technique for estimating data-driven statistical price indices.

2.4 Statistical real estate price indices

2.4.1 Hedonic approach

The hedonic approach is used in many areas where prices of heterogeneous observable objects are scrutinized. It explicitly controls for heterogeneity in a given sample. The approach is also known as hedonic regression in the literature because it relates observed prices to characteristics by regression methods. Hedonic regressions have been used for analyzing prices of asparagus, automobiles, and computers among others (Griliches; 1971; Triplett; 1987; Berndt; 1991; Berndt et al.; 1995). It is a prominent tool in analyzing real estate prices, and it is often used for calculating real estate price indices, see the surveys of Sheppard (1999) and Malpezzi (2002).

It is the core assumption of the hedonic approach that prices are given by a function of property-specific characteristics. The location model presented in Section 2.2 can serve as a motivation for this assumption. According to this model, prices of heterogeneous properties are influenced by common market factors and the location and size of the property. Closer to reality, hedonic regression includes much more characteristics.

Let $P_{n,t}$ denote the transaction price of house n that is sold in period t and let $x_{n,t}$ denote its vector of characteristics, which contains—among other things—the age of the building, the size of the lot, the size of the floor space, location information. Then the price of the property is related to its characteristics by

$$P_{n,t} = P_t(x_{n,t}) \ .$$

Here, $P_t(\cdot)$ is the so-called price function which gives the price of property n in period t as a function of its characteristics $x_{n,t}$.

In empirical applications, time effects are often captured by dummy variables for the respective period and the price function is related to transformed prices, so that

$$T(P_{n,t}) = \beta_t + P(x_{n,t}) \ ,$$

where $T(\cdot)$ is a transformation function for the price $P_{n,t}$ and β_t is a general market trend for real estate, which captures common market factors. A price index for a standardized property with characteristics x_s is then given as

$$I_{0t} = \frac{T^{-1}\{\beta_t + P(x_s)\}}{T^{-1}\{\beta_0 + P(x_s)\}}$$

In reality, β_t and the price function are unknown and have to be estimated. Thus, an applied hedonic equation has the form

$$T(P_{n,t}) = \beta_t + P(x_{n,t}) + \varepsilon_{n,t} \ , \tag{2.34}$$

where $\varepsilon_{n,t}$ is a disturbance term, capturing unobservable, idiosyncratic influences on the price of a house. Calculating a price index for a class of real estate requires to specify a functional form for the price function and to estimate its unknown coefficients.

Functional form of hedonic equations

Economic theory should guide the choice of functional forms for hedonic equations. But, alas, economic theory does provides only rather limited guidance.

In the classical theory of consumer behavior, households derive utility $u(z)$ by consuming a bundle of market goods z. In hedonic models, consumers derive utility $u(x)$ from the intrinsic characteristics of market goods, where every good i is completely described by its characteristics. The spirit of these models is that market goods are transformed into ultimate consumable characteristics (Lancaster; 1966; Rosen; 1974; Muellbauer; 1974). In his classic paper Lancaster (1966) assumes that market goods z are divisible and can be combined with a linear consumption technology. Given the bundle of goods z and given a linear consumption technology B, the quantity of characteristics is $x = Bz$. Here, B consists of vectors $b_i = [b_{i1}, \ldots, b_{iK}]^\top$ which contain the characteristics contained in one unit of good z_i. It is important to emphasize that a vector of characteristics can be replicated by different combinations of goods (Gravelle and Rees; 1992, Chap. 5 B). If cost-free repacking is possible and a x can be obtained by different combinations of traded goods, then the price function $P(\cdot)$ is linear in the characteristics (Rosen; 1974). However, the assumption of repackaging is implausible for properties: two buildings with different locations cannot be combined into one building with characteristics that are the sum of each other.

Different to Lancaster, Rosen (1974) uses a model where repackaging is impossible. Rosen shows that the hedonic price function $P(x)$ represents a joint envelope of household bid functions and producer offer functions, where only linear forms can be ruled out. However, he also shows that the functional form of $P(\cdot)$ cannot be determined on theoretical grounds.

Rosen's paper is mentioned in nearly every paper that uses the hedonic approach for evaluating the behavior of house prices because it gives a justification for pragmatically choosing a functional form for $P(x)$, see for example Palmquist (1980); Cropper et al. (1988); Engle et al. (1985); Quigg (1993); Meese and Wallace (1997); Berger et al. (2000).

Most hedonic studies use parametric specifications and tend to rely on only few linear and log-linear forms (Sheppard; 1999). Functions that are more flexible are obtained by using the Box-Cox transformations

$$T_\lambda(Z) = \begin{cases} \dfrac{Z^\lambda - 1}{\lambda} & \text{when } \lambda \neq 0 \ , \\ \ln Z & \text{when } \lambda = 0 \ , \end{cases}$$

where Z is the variable to be transformed and λ is the transformation parameter (Halvorsen and Pollakowski; 1981). λ may be known or—if unknown—it has to be estimated (Davidson and MacKinnon; 1993, Chap. 14). The hedonic regression equations are

$$T_{\lambda_1}(P_{n,t}) = \beta_t + \sum_{k=1}^{K} \beta_{k,x} T_{\lambda_2}(X_{k,n,t}) + \varepsilon_{n,t} \ , \tag{2.35}$$

where $\varepsilon_{n,t}$ is a disturbance term, β_t is a time dummy that captures the general market level in period t and $\beta_{k,x}$ is the implicit price for the transformed characteristic k. If $\lambda_1 = \lambda_2 = 1$, the above equation reduces to the linear model. If $\lambda_1 \to 0$ and $\lambda_2 \to 0$, then the above equation approaches the log-linear model, and if $\lambda_1 \to 0$ and $\lambda_2 = 1$ it approaches the semi-log model.

Hedonic regression at work

Reported price indices should control explicitly for the heterogeneity of observed transaction prices. Practitioners sometimes report prices per square meter and thus try to adjust partly for the heterogeneity. But this may be misleading if other characteristics influence the price. We will clarify this point in this section. For a convenient exposition, we suppose that the hedonic equation (2.34) is linear in log prices and in (possibly transformed) characteristics with

$$p_{n,t} = \beta_t + x_{n,t}\beta_x + \varepsilon_{n,t} \ , \tag{2.36}$$

where $p_{n,t} \stackrel{\text{def}}{=} \ln P_{n,t}$ and the row vector $x_{n,t}$ contains the characteristics. We assume that the disturbances fulfill $\varepsilon_{n,t} \sim (0, \sigma_\varepsilon^2)$ for all n and t. All disturbances are pairwise uncorrelated. The market level of log prices is β_t. The implicit prices for the characteristics are collected in the $(K \times 1)$ vector β_x.

We have T time periods, where we observe $N_t \geqslant 1$ sales of heterogeneous properties per period. The total number of sales is

$$N \stackrel{\text{def}}{=} \sum_{t=1}^{T} N_t \ . \tag{2.37}$$

with $N > T + K$.

For easier handling, we write the system of equations in matrix form. Let p_t denote the $N_t \times 1$ vector with all price observations for period t, 1_{N_t} a unit vector with N_t rows and 0_{N_t} a vector of zeros with N_t rows. The $N_t \times K$ matrix X_t contains the characteristics of all objects sold in t. The disturbances are comprised in the vectors ε_t.

Defining

$$\beta_c \stackrel{\text{def}}{=} [\beta_1, \dots, \beta_T]^\top \ ,$$

which collects the T common price components, the system of equations is

$$\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_T \end{bmatrix} = \begin{bmatrix} 1_{N_1} & 0_{N_1} & \dots & 0_{N_1} & X_1 \\ 0_{N_2} & 1_{N_2} & \dots & 0_{N_2} & X_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{N_T} & 0_{N_T} & \dots & 1_{N_T} & X_T \end{bmatrix} \begin{bmatrix} \beta_c \\ \beta_x \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix} \ . \tag{2.38}$$

We write this system compactly as

$$p = Z\beta + \varepsilon \tag{2.39}$$

with

$$\beta \stackrel{\text{def}}{=} \begin{bmatrix} \beta_c \\ \beta_x \end{bmatrix}$$

and

$$Z \stackrel{\text{def}}{=} \begin{bmatrix} D & X \end{bmatrix} \ , \tag{2.40}$$

where the $N \times T$ matrix D

$$D \stackrel{\text{def}}{=} \begin{bmatrix} 1_{N_1} & 0_{N_1} & \dots & 0_{N_1} \\ 0_{N_2} & 1_{N_2} & \dots & 0_{N_2} \\ \vdots & \vdots & \vdots & \\ 0_{N_T} & 0_{N_T} & \dots & 1_{N_T} \end{bmatrix}$$

contains all entries for the common price component and the $N \times K$ matrix X is just the matrix with all matrices of characteristics stacked on each other, see (2.38).

Given that the explanatory variables are not linearly dependent, i.e. $\operatorname{rank}(Z) = T + K$, and can be treated as fixed, OLS is the best linear unbiased estimator (BLUE) for β (Davidson and MacKinnon; 1993, pp. 159). The estimator for the common price component and implicit prices is

$$b = (Z^\top Z)^{-1} Z^\top p \,. \tag{2.41}$$

Whereas the first T components of b are unbiased for β_c, average prices may be biased. To see this, we use the following $(T \times N)$ matrix

$$A \stackrel{\text{def}}{=} (D^\top D)^{-1} D^\top = \begin{bmatrix} 1/N_1 & 0 & \dots & 0 \\ 0 & 1/N_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 1/N_T \end{bmatrix} \begin{bmatrix} 1_{N_1}^\top & 0_{N_2}^\top & \dots & 0_{N_T}^\top \\ 0_{N_1}^\top & 1_{N_2}^\top & \dots & 0_{N_T}^\top \\ \vdots & & & \vdots \\ 0_{N_1}^\top & 0_{N_2}^\top & \dots & 1_{N_T}^\top \end{bmatrix} ,$$

that generates the averages per period for post-multiplied matrices p and X.

RESULT 2.6 (Average prices may be misleading). *Using the partition of the matrix Z, it is easy to show that the expectation of average prices as an estimator for the common price component is given by*

$$\mathcal{E}[Ap] = \beta_c + AX\beta_x \,.$$

Average prices are an unbiased estimator for the common price components only if $\beta_x = 0$ or if the evaluated characteristics per period sum to zero, that is, if $AX\beta_x = 0$. Otherwise, the per period bias is proportional to evaluated average characteristics.

This confines the well-known conjecture that average prices are biased estimators because they do not attempt to control for heterogeneity of houses sold during different periods (Meese and Wallace; 1997; Case and Quigley; 1991).

Hedonic regression controls explicitly for the influence of the characteristics by correcting average prices with respect to the empirical covariances between prices and characteristics. To derive this result, we decompose the estimator b into the estimator for the common price components, i.e., the first T entries in b, which we denote with b_c and the estimator b_x for the implicit prices.

First of all, we note that

$$Z^\top Z = \begin{bmatrix} D^\top D & D^\top X \\ X^\top D & X^\top X \end{bmatrix} .$$

Using a well-known matrix inversion lemma (Sydsæter et al.; 2000, 19.48), we obtain for $(Z^\top Z)^{-1}$ the following partition

$$\begin{bmatrix} \{(D^\top D)^{-1} + AX\Sigma^{-1}(AX)^\top\} & -AX\Sigma^{-1} \\ -\Sigma^{-1}(AX)^\top & \Sigma^{-1} \end{bmatrix} \tag{2.42}$$

with

$$\Sigma = (MX)^\top (MX)$$

and

$$M = [I_N - D(D^\top D)^{-1} D^\top] .$$

I_N is a $(N \times N)$ identity matrix. It is easy to check that the $(N \times N)$ matrix M is a singular symmetric idempotent matrix with $D^\top M = [0_T, \dots, 0_T]$ and

$$\begin{aligned} \operatorname{rank}(M) &= \operatorname{tr}(M) \\ &= N - T , \end{aligned}$$

where $\operatorname{rank}(M)$ gives the rank of matrix M and $\operatorname{tr}(M)$ the trace of matrix M, i.e., the sum of its diagonal elements. If we multiply M with X, we obtain a $N \times K$ matrix

$$\begin{bmatrix} (I_{N_1} - \frac{1}{N_1} 1_{N_1} 1_{N_1}^\top) X_1 \\ (I_{N_2} - \frac{1}{N_2} 1_{N_2} 1_{N_2}^\top) X_2 \\ \vdots \\ (I_{N_T} - \frac{1}{N_T} 1_{N_T} 1_{N_T}^\top) X_T \end{bmatrix}$$

that gives the deviations of every characteristic from the average of that characteristic per period. Thus, the diagonal elements of Σ are given by

$$\begin{aligned} \Sigma_{kk} &= \sum_{t=1}^{T} \sum_{n=1}^{N_t} (x_{k,n,t} - \bar{x}_{k,t})^2 \\ &= \sum_{t=1}^{T} N_t \widehat{\mathcal{V}}[x_{k,t}] \end{aligned}$$

and the off-diagonal elements by

$$\begin{aligned}\Sigma_{kj} &= \sum_{t=1}^{T}\sum_{n=1}^{N_t}(x_{k,n,t}-\bar{x}_{k,t})(x_{j,n,t}-\bar{x}_{j,t}) \\ &= \sum_{t=1}^{T} N_t\widehat{\mathcal{C}}(x_{k,t},x_{j,t}) \ ,\end{aligned}$$

Here, $\mathcal{V}[\cdot]$ denotes the variance and $\mathcal{C}[\cdot]$ the covariance. Divide now Σ by N to obtain the neatly interpretable expressions

$$\frac{1}{N}\Sigma_{kk} = \sum_{t=1}^{T} w_t\widehat{\mathcal{V}}[x_{k,t}]$$

and

$$\frac{1}{N}\Sigma_{kj} = \sum_{t=1}^{T} w_t\widehat{\mathcal{C}}[x_{k,t},x_{j,t}] \ .$$

with $w_t \overset{\text{def}}{=} N_t/N$. The elements of Σ are proportional to the weighted covariances of the different characteristics. The weights are given by the relative proportion of observations.

RESULT 2.7 (Hedonic regression adjusts average prices). *By post-multiplying the upper T rows of $(Z^\top Z)^{-1}$ given in equation (2.42) with $Z^\top$, we obtain for the first T elements of b*

$$b_c = Ap - AX\Sigma^{-1}X^\top Mp \ . \tag{2.43}$$

Using the lower partition of $(Z^\top Z)^{-1}$, it is easy to check that

$$b_x = \Sigma^{-1}(MX)^\top Mp \ .$$

Thus, the estimates from OLS and per period average prices Ap are only equal if $b_x = 0$ or if $AXb_x = 0$.

The elements of the $T \times 1$ vector $(MX)^\top Mp$ are

$$\sum_{t=1}^{T} N_t\widehat{\mathcal{C}}[x_{i,t},p_t]$$

and thus weighted sums of covariances between the characteristics and the prices. This fact allows an easy interpretation of b_c: if the characteristics and the prices are uncorrelated, the OLS estimates of the common price components are just the average prices. If prices and characteristics are correlated, then the estimates correct for these influences. The second expression in (2.43) is a weighted per-period average of the characteristics that is subtracted from the average price per period. The weights are given by

$$\Sigma^{-1}(MX)^{\top}Mp$$

and thus the weighted covariances of characteristics and prices. The weights are the overall variation in the characteristics.

If the primary goal is a price index, then a further interesting result can be offered. We obtain a log price index for real estate if we subtract the common price component of the base period from the common price components of the other periods. Using the matrix

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & 0 & \dots & 0 \\ -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix} ,$$

we obtain the index series calculated with per period average prices as

$$\ln I_{avg} = BAp \; .$$

It is easy to check that

$$\mathcal{E}[\ln I_{avg}] = B\beta_c + BAX\beta_x \; .$$

RESULT 2.8 (Matched index). *If we observe identical houses in every period, then*

$$BAX = 0 \tag{2.44}$$

and average log prices are an unbiased estimator for the price index, even if $AX\beta_x \neq 0$. *Under this assumption, it is also easy to see that price indexes calculated with OLS are identical to* $\ln I_{avg}$. *The OLS indexes are given by*

$$\ln I_{ols} = Bb_c \; . \tag{2.45}$$

Plugging (2.43) into (2.45) delivers

$$\ln I_{ols} = BAp - BAXb_x \ .$$

and we obtain immediately the result with (2.44). If we use the antilog, we obtain

$$I_{1t} = \prod_{n}^{N_t} \left(\frac{P_{n,t}}{P_{n,1}} \right)^{1/N_t}$$

for $t > 1$, N_t *constant for all periods, and* $I_{11} = 1$*. Given that we observe identical objects in every period and that the hedonic equation is linear in log prices and characteristics, the Jevons index is a reasonable price index.*

2.4.2 Repeat sales technique

The repeat sales technique uses the above result that one can control for heterogeneity in a log-linear hedonic regression with sales observations of identical objects, where the objects have to be observed at least two times. It is not necessary that the observations are in consecutive periods.

The repeat sales technique was first proposed for calculation of real estate price indices by Bailey et al. (1963). The approach controls for characteristics by using repeat sales and assuming that the characteristics remain constant between sales. Using our simple model, where the price of house n is given by (2.36)

$$p_{n,t} = \beta_t + x_{n,t}\beta_x + \varepsilon_{n,t} \ .$$

and assuming that the same house is sold a second time in period $t' > t$, then given unchanged characteristics, we obtain

$$\Delta p_{n,tt'} = \beta_{t'} - \beta_t + u_{n,tt'} \tag{2.46}$$

with

$$\begin{aligned} \Delta p_{n,tt'} &\stackrel{\text{def}}{=} p_{n,t'} - p_{n,t} \\ u_{n,tt'} &\stackrel{\text{def}}{=} \varepsilon_{n,t'} - \varepsilon_{n,t} \ . \end{aligned}$$

This simple equation states the very idea of the repeat sales approach: given identical objects, the evaluated characteristics cancel each other out.

Given unchanged evaluated characteristics, the repeat sales approach has the attractive feature that only transaction prices are necessary to estimate an real estate price index. Let us normalize the first common price component to zero. Given the above equation, this is achieved by subtracting β_1 from every common price component

$$\beta_t^{norm} \stackrel{\text{def}}{=} \beta_t - \beta_1 \ .$$

To be aware of this necessary normalization (see below), we use the superscript *norm*. Using the $\{1 \times (T-1)\}$ vector $R_{n,tt'}$ that has a -1 in column $t-1$ and a 1 in column $t'-1$ and zeros otherwise, we can write the above equation compactly as

$$\Delta p_{n,tt'} = R_{n,tt'} \beta^{norm} + u_{n,tt'} \ ,$$

where the vector β^{norm} comprises the $(T-1)$ normalized common price components. Stacking the price relatives, the corresponding $R_{n,tt'}$ vectors and the noise terms into the vector Δp, the matrix R and the vector u, we obtain the system

$$\Delta p = R\beta^{norm} + u \ . \tag{2.47}$$

If a house is not sold in the first period, then we have $R_{n,tt'} 1_{T-1} = 0$ for $t > 1$. If a house is sold in the first period, $R_{n,1t'}$ has a 1 at column $t'-1$ and zeros otherwise. Now it is immediately obvious why some normalization is needed: without normalization, the vectors $R_{n,1t'}$ would have the dimension $(1 \times T)$ and a -1 as first entry for all houses that are sold in the first period. Multiplying that vector with a $(T \times 1)$ unit vector gives 0 and thus the rank of matrix R would be smaller than T. In that case, $R^\top R$ is singular and a basic assumption for applying OLS is violated. The necessary normalization is intuitively understandable if one recalls that we work with price ratios. Any absolute price levels cancels out if we calculate such ratios.

In addition to the normalization of the common price components, two further necessary and sufficient conditions on the data must be fulfilled to guarantee that $R^\top R$ is of full rank (Webb; 1988). The first condition is that at least one transaction must be observed in every period. The second condition is that for every period t there must exist at least one house bought in a period before t and sold in a period after t. Webb calls data that satisfy both conditions connected, see also Searle (1971, 7.4.4). Intuitively, without observations in period t, it is impossible to estimate β_t^{norm}. However, even if we observe sales in t, but only first sales and all houses that were sold the first time before period t were also sold the second time before t, then we have no observation that connects period t with the periods before t. In that case, it is impossible to extract the price change between $t-1$ and t.

In the model given above, the common price component β_t^{norm} for $t \in \{2, \dots, T\}$ gives the total relative change of the common house price component from period 1 to period t. Quite intuitively, the common price change between consecutive periods is $\beta_t^{norm} - \beta_{t-1}^{norm}$. The above presented model can be easily transformed into a model for per period relative price changes. We use the transformation matrix

$$F = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

and define

$$\tilde{\beta}^{norm} \stackrel{\text{def}}{=} F\beta^{norm} \, .$$

The vector $\tilde{\beta}^{norm}$ gives the relative per period changes in real estate prices. It is easy to prove that F is just the inverse of the lower triangular matrix with 1 as entries. We have

$$\Delta p = \tilde{R}\tilde{\beta}^{norm} + u \, ,$$

where every row of $\tilde{R} \stackrel{\text{def}}{=} RF^{-1}$ has a 1 in column t if the owner still holds the house in that period and zeros otherwise.

Does it matter which formulation of the relative price change is used? It is clear that this modelling decision should have no influence on the results. Given the OLS estimator we obtain by using elementary matrix operations

$$\begin{aligned} \tilde{b}^{norm} &= (\tilde{R}^\top \tilde{R})^{-1} \tilde{R}^\top \Delta p \\ &= F(R^\top R)^{-1} R^\top \Delta p \\ &= Fb^{norm} \end{aligned}$$

and the result is not influenced by the choice of the two different modelling strategies.

There are several extension of the above presented repeat sales technique. Webb (1988) assumes that the noise term is the sum of per period disturbance terms, so that the disturbances in the repeat sales equation are heteroscedastic and variance are proportional to the number of periods between sales. Further, Case and Shiller (1987, 1989) assume that a disturbance term is due to unusual circumstances at the purchase date. Under both assumptions, Generalized Least Squares (GLS) is used for estimation. Furthermore, Shiller (1993a,b) extends the repeat sales model by including observed changes in characteristics and still assuming that unobserved characteristics are unchanged. In a similar fashion Case and Quigley (1991), Hill et al. (1997), and Englund et al. (1998) combine single sales observations and repeat sales observations all in one joint estimation, where improvements in repeatedly sold properties are explicitly considered. Such hybrid techniques are an interesting strategy because they combine the classical hedonic technique for single-sales with the repeat sales technique.

2.4.3 Which technique should be used?

Given the different techniques for estimating the common market trend, one may ask which one should be used? One can answer this question very pragmatically, because it depends mainly on the data at hand (Pollakowski; 1995). Given privacy restrictions, it may be impossible to figure out sales of identical properties. Even if repeat sales can be figured out, it may be the case that the number of these properties is too small to derive accurate estimates.

There are several studies that compare empirically the different techniques. Goetzmann (1992) compares several repeat sales estimators with different assumptions on the disturbance term in a simulation study by randomly sampling stocks from the New York Stock Exchange and censoring the daily data to mimic repeat sales data. He finds that all repeat sales estimators perform well when the number of repeat sales is large. Case and Szymanoski (1995) compare for a hedonic model with single-sales, a repeat sales and a hybrid model the estimated standard deviations, the width of a 95 percent confidence interval around the predicted price, and the R^2 and found the best method is the hybrid model. The results of Meese and Wallace (1997) are similar and show that hedonic models with single-sales data are favorable because the repeat sales indices are quite sensitive to the small sample problem. Using several prediction error measures, Clapp and Giaccotto (2002) derive that the hedonic technique with single-sales data dominates the repeat sales technique. These accuracy studies show that the classical repeat sales technique is not necessarily better than the classical hedonic technique with single-sales data.

There is another important aspect of the hedonic approach applied to real estate prices. The approach lacks of an intuitive economic explanation for the common price component and the implicit prices. The next chapter presents a model which delivers a motivation for the price component and the implicit prices. It is closely connected to the classical hedonic technique.

Chapter 3

A State Space Model for Berlin Single-Family House Prices

3.1 Introduction

For many households, owner-occupied housing is not only a place to live but also the single most important asset in their portfolio. Indeed, in most industrialized countries real estate is the greatest component of private households' wealth. As a consequence, the value of their home has a major impact on households' consumption and savings opportunities (Case et al.; 2001). House prices are therefore of great interest to actual and potential home owners but also to real estate developers, banks, policy makers or, in short, the general public.

Observed house prices show considerable variation in both the cross-section and time-series dimensions. The cross-section variation derives from the heterogeneity of houses, which—at any given date in any city—usually vary widely in age, size, location and many other characteristics. The time-series variation is driven by the changing conditions in the housing market as a whole and represents the component of prices that is common to all houses in the same market. An empirical model of house prices has to account for these two sources of variation in house prices.

Hedonic regression, the standard approach for analyzing individual house price data, is predicated on including observed house characteristics as variables explaining house prices (Case and Quigley; 1991; Hill et al.; 1997; Dombrow et al.; 1997; Shiller; 1993a,b). It is therefore well suited for taking account of the cross-section variation of house prices. Section 2.4 of the previous chapter gives a discussion on hedonic regression. In comparison, the treatment of the over-time variation of the common price component is usually rather unsophisticated in hedonic regression: time dummies are used to capture the law of motion of the general price trend. The movement of their coefficients are not required to follow a particular model of univariate time series and are not allowed to depend on observable variables that may help to explain the behavior of the common price component. Hence, the standard model offers little potential for explaining and predicting the time series variation of house prices.

In this chapter, we analyze the prices of single family homes in Berlin, Germany, with a statistical model designed to overcome this weakness of the conventional hedonic approach. Our model, building on the work by Engle et al. (1985) and Schwann (1998), augments hedonic regression with an explicit statistical model for the common component of prices. Moreover, we address two other often-criticized properties of the hedonic methodology: its loose connection to economic theory and its ad-hoc functional form. We derive our hedonic equation from the present value model of asset prices, thus providing theoretical motivation and aiding interpretation of the hedonic model and the common price component. Still, the present value model does not completely pin down the functional form of the hedonic equation. We use a model selection procedure that is based on a cross-validation criterion to choose between various possible transformations of the continuous explanatory variables, see Bunke et al. (1999).

In addition to employing a statistical model that attempts to improve upon the conventional hedonic approach in several respects, by using data from Berlin, Germany, this chapter also contributes to the surprisingly small literature devoted to the real estate market in Europe's biggest economy. Our estimates of the coefficients of the hedonic equation provide plausible and easily interpretable figures of the premiums or rebates that different house characteristics command. The estimated process of the common price component is highly persistent and sluggish.

The remainder of this chapter is organized as follows: we derive the hedonic equation in Section 3.2 with the help of the present value model. In Section 3.3 the empirical model is written as a state space model and the estimation strategy is laid out. It basically consists of combining the Kalman filter and the EM algorithm to get estimates of the unknowns in our model. Section 3.4.1 describes the data. The empirical results are presented in Section 3.4.2. Section 3.5 concludes.

3.2 Value of a house

We use the asset market approach of Poterba (1984, 1991) and Miles (1994) to motivate the statistical model used in the empirical part of the paper. The asset market approach is presented in Section 2.2. According to this approach, the value of a house equals the expected discounted value of its future stream of housing services, irrespective if the house is rented out or owner-occupied. For rented properties, this present value has an obvious equivalent in the expected discounted sum of net operating rents. For owner-occupied homes, there is no direct empirical analog, as owner-occupants do not explicitly receive (as owners) or pay (as occupants) rent. (Potential) owner-occupants, however, have the option to satisfy their demand for accommodation by renting rather than owning a house or an apartment. Market rents are, therefore, the opportunity costs of owning a house. Many (would-be) owner-occupiers explicitly 'face' these costs when they apply for a mortgage credit: for calculating the maximal debt service, banks add rents to the current disposable income of the applicant.

However, a prerequisite for this link is an extensive rental market. In Berlin, only 10.4% of the 1.9 Million residential dwellings are owner-occupied (Senatsverwaltung für Stadtentwicklung und Investitionsbank Berlin; 2002), thus indicating that the rental market delivers interesting opportunities for potential house owners. 22.9% of all rental dwellings in Berlin are social housing, where the rents are regulated at a level well below market rents. However, social housing is restricted to low income households, and is not obtainable for households with income high enough to afford an own house. Moreover, information on rents is easily obtained from the rent-surveys local authorities are obliged to publish regularly. Those rent-surveys are required because the German Rent Law curtails rent increases within three years (to 30% in the period of interest) and restricts the absolute level above by surveyed rents from comparable dwellings. The intention of the regulation is to prevent rents in old contracts from lagging too much behind new contracts and to guarantee for a low dispersion of rents for comparable properties.

Owner-occupiers in Berlin are therefore well aware of the opportunity costs of owning their home. In reality owner-occupied houses may have characteristics non-comparable with rental dwellings and may be exposed to favorable tax treatment. In this case, however, households will derive imputed rents by adding valued non-comparable characteristics from owner-occupied housing to market rents.

The asset market approach states that a (potential) home owner values a house according to its present value

$$V_{n,t} = \sum_{j=1}^{\infty} \mathcal{E}_t \left[\frac{D_{n,t+j}}{\prod_{i=1}^{j}(1+R_{t+i})} \right] . \tag{3.1}$$

Here, $V_{n,t}$ denotes the market value of house n sold in period t and $D_{n,t+j}$ denotes the net operating rent, i.e. (imputed) rents without maintenance costs and depreciation. $\mathcal{E}_t[\cdot]$ is a shorthand for the expectation taken conditional on the information available at time t. R_{t+i} are time-varying discount rates. Observed transaction prices may deviate from the market value due to unusual circumstances and idiosyncratic influences during the price negotiations. Reformulating (3.1) gives

$$V_{n,t} = \mathcal{E}_t \left[\frac{D_{n,t+1} + V_{n,t+1}}{1+R_{t+1}} \right] .$$

Market values can be likewise seen as based on expected capital gains plus the expected one-period flow of housing services (Miles; 1994, Chap. 2). Under the assumptions of the Gordon growth model—discount rates R and expected rent growth rate G are constant and $R > G$—the present value (3.1) boils down to

$$V_{n,t} = \frac{D_{n,t}}{\theta} \tag{3.2}$$

with $\theta \stackrel{\text{def}}{=} (R-G)/(1+G)$. Such average implied cap rates θ are often reported as measures of return in real estate investments. Using price to net rent multipliers provided by the Federal Organization of Estate Agents and Property Management, the average yearly implied cap rate for apartment houses in Berlin was 6.9% between 1989-1999 for properties built after 1948 and 7.5% for older buildings (Ring Deutscher Makler; 1989-1999).

Instead of working with equation (3.1) directly, we use a log-linearized version of it (Campbell et al.; 1997; Cochrane; 2001). Let r_t denote the log of one plus the return rate and $d_{n,t}$ the log of the net operating rent. The first order approximation for the log market value is

$$v_{n,t} = \frac{k}{1-\rho} + d_{n,t} + \sum_{j=0}^{\infty} \rho^j \Big(\mathcal{E}_t[\Delta d_{n,t+1+j}] - \mathcal{E}_t[r_{t+1+j}] \Big) . \tag{3.3}$$

with $k = \ln(1+\theta) - \theta \ln \theta/(1+\theta)$, $\rho \stackrel{\text{def}}{=} 1/(1+\theta)$ and the long-run discount rate θ. We have $\theta > 0$ and $\rho < 1$. It is easy to see from (3.3) that ρ

can be interpreted as a discount factor in the linearized model, where the discount rate is given by θ. Δ denotes the difference operator, so that $\Delta d_{n,t}$ gives approximately the growth rate of the net operating rents. Under the assumptions of the Gordon growth model (3.3) reduces to $v_{n,t} = d_{n,t} - \ln\theta$ and the approximation of the log market value is exact.

Given our discussion on the rich set of rental opportunities in Berlin, we model the net operating rents with a hedonic function of the house characteristics which is proportional to the general city-wide rent level. Let d_t^0 denotes the log of Berlin's rent index. We will refer to the notional object that corresponds to this index as the reference dwelling. Then we have

$$d_{n,t} = \delta + d_t^0 + x_{n,t}\beta + \varepsilon_{n,t} \tag{3.4}$$

where $\varepsilon_{n,t}$ is white noise and the row vector $x_{n,t}$ comprises the (possibly transformed) characteristics for house n such as its age or its floor size. The constant δ absorbs the normalization of the rent index and the characteristics of the reference dwelling weighted by the implicit prices β. There are several non-comparable features of owner-occupied housing, which are also captured by δ. One feature is house-ownership per se, which may command a premium, because it gives the owner the right to model the object in accordance to his own taste. Another feature may be lower maintenance cost due to the absence of the principal agent problem between lessor and lessee (Homburg; 1994). A unobservable renter will handle a dwelling with less care than owners would do and these extra cost should be absent in imputed rents. Ownership also means incurring costs like wealth or property taxes, and it depends on the incidence if changes in tax rates are captured by the city-wide rent index. Changes in the promotion of owner-occupied housing, however, have to be controlled explicitly.

The expected change in the imputed rent of house n is equal to the relative change of the general rent level, so that substitution of (3.4) into (3.3) gives

$$p_{n,t} = \kappa + p_t^0 - \sum_{j=0}^{\infty} \rho^j \mathcal{E}_t[r_{t+1+j}] + x_{n,t}\beta + \varepsilon_{n,t} \ , \tag{3.5}$$

where κ absorbs all constants. The log market value is replaced by the log price, where unusual circumstances will be incorporated in $x_{n,t}$ and idiosyncratic influences during the negotiations are considered in the noise term $\varepsilon_{n,t}$. Moreover,

$$p_t^0 \stackrel{\text{def}}{=} d_t^0 + \sum_{j=0}^{\infty} \rho^j \mathcal{E}_t[\Delta d_{t+1+j}^0] \tag{3.6}$$

is up to a constant equal to the value of the reference dwelling. We will designate p_t^0 as reference value.

Finally, we specify the process of the return rate as

$$r_{t+1+j} = \phi r_{t+j} + (1-\phi)r^* + s_{t+j}\gamma + \nu_{t+1+j} \ . \tag{3.7}$$

The random component ν_t is white noise. Or specification copes the empirical fact that house price returns are autocorrelated (Englund and Ioannides; 1997). The required return depends on its lagged values and on the long-run rate r^*. Furthermore, the return is influenced by observable shocks of market indicators, which are collected in the row vector s_t. We obtain after some manipulations that

$$\mathcal{E}_t[r_{t+1+j}] = r^* + \phi^j \mathcal{E}_t[r_{t+1} - r^*] \quad \text{for} \quad j \geqslant 0 \ . \tag{3.8}$$

It is easy to see that the long run required rate is equal to r^* for $|\phi| < 1$. If we substitute (3.8) into the price equation (3.5), define $r_{t+1}^e \overset{\text{def}}{=} \mathcal{E}_t[r_{t+1} - r^*]$ and assume $|\phi| < 1/\rho$ we get

$$p_{n,t} = \kappa + p_t^0 - \frac{1}{1-\rho\phi} r_{t+1}^e + x_{n,t}\beta + \varepsilon_{n,t} \ , \tag{3.9}$$

where, once again, all constants are absorbed by κ. The expected changes in the return rate, r^e, are unobservable. Rewriting (3.7) in deviation form for $j = 0$, taking expectations at t and using $r_t - r^* = r_t^e + \nu_t$, we derive

$$r_{t+1}^e = \phi r_t^e + s_t\gamma + \xi_t \tag{3.10}$$

with $\xi_t \overset{\text{def}}{=} \phi\nu_t$. Let ψ denote $1/(1-\rho\phi)$ and multiply the above equation with this term, we eventually arrive at the following two equations system

$$\Delta^0 p_{n,t} = \kappa - r_{\psi,t+1}^e + x_{n,t}\beta + \varepsilon_{n,t} \tag{3.11a}$$

and

$$r_{\psi,t+1}^e = \phi r_{\psi,t}^e + s_t\gamma_\psi + \xi_{\psi,t} \ . \tag{3.11b}$$

Here, $\Delta^0 p_{n,t}$ denotes $p_{n,t} - p_t^0$ and the subscript in the return equation indicates the transformation of (3.10) using ψ. The dependent variable of the price equation $\Delta^0 p_{n,t}$ is the difference between the observable sale price $p_{n,t}$ and the value of the reference dwelling, p_t^0. Since p_t^0 is not directly observable, we have to estimate it. Details are given in Appendix 3.6.2.

The two-equation system (3.11) differs in several important respects from a standard hedonic model of house prices. By using asset pricing theory to derive (3.11a), we can give a well-grounded economic interpretation of this equation: the deviation between the current price of house n and the reference value is a function of the characteristics of the object, and the cumulated effect of the current return rate deviation. Similarly, we can give economic content to the component of prices that is common to all houses sold in the same period: it coincides with the return rate deviation in our model. Moreover, equation (3.11b) poses a model for the law of motion of this unobservable return rate. This equation says that the cumulated return deviations are influenced by their previous value and by observable market indicators. Moreover, they are influenced by other unobservable, unsystematic shocks to expected returns.

This is in contrast to a one-equation hedonic price equation where the common price component is captured by time dummies whose coefficients are not restricted in the estimation process. While the two equation treatment is less flexible in this respect it offers the possibility to explain and forecast the movement of the common price component. Despite postulating one equation with an unobserved dependent variable, the unknown parameters in both equations of (3.11) can be estimated by putting the system in state space form and applying the Kalman filter to infer r^e_ψ from the observed movements in individual house prices. This approach is also able to handle the fact that the number of observations varies between periods. It is described in the following section.

3.3 State space model and estimation algorithm

Before turning to the state space version of the two-equation model (3.11) we describe the model selection procedures used to determine the precise specification of the price equation. The theory of hedonics (Rosen; 1974; Palmquist; 1980) does not suggest a particular functional form for the relationship (3.4) between the net operating rent $d_{n,t}$ and the observable characteristics of a house, collected in the vector $x_{n,t}$ (and, by virtue of (3.3) and (3.5), for the relationship between $p_{n,t}$ and $x_{n,t}$). In our data, most components of $x_{n,t}$ are dummy variables representing various qualitative characteristics, which naturally enter the model in a linear way. Regarding the continuous components of $x_{n,t}$, we rely on statistical model selection criteria to pin down the functional form of the relationship between these variables and $\Delta^0 p_{n,t}$. For each continuous regressor, we choose from a set of Box-Cox type transformations $T_\lambda(\cdot)$ indexed by the parameter λ. For the continuous explanatory variables of the price equation we consider the following transformations:

$$T_\lambda(x) = \begin{cases} \lambda^{-1}\left[\{s^{-1}(x+a_\lambda)\}^\lambda - 1\right] & \text{for } \lambda \in \Lambda, \\ \ln\{s^{-1}(x+a_0)\} & \text{for } \lambda = 0 \end{cases} \tag{3.12}$$

with $\Lambda = \{-2, -1, -0.5, 0.5, 1, 2\}$. Here x denotes any of the continuous explanatory variables, a_λ is a constant depending on λ, s is the sample standard deviation of variable x and λ is the parameter that determines the transformation. A particular value of λ implies a value of the constant a_λ. These constants are computed according to the suggestions made in Bunke et al. (1999) and aim to make, for any given λ, the transformation as nonlinear as possible. Therefore, different to the Box-Cox transformations for equation (2.35) in Section 2.4 of the previous chapter, the continuous variables may be transformed by different λs.

For finding the optimal values of λ we do not use the state space model framework described in the following section but rather adopt the standard one-equation hedonic regression approach with time dummies. More precisely, we work with the price equation

$$\Delta^0 p_{n,t} = I_t + x_{c,n,t}\beta_c + x_{d,n,t}\beta_d + \varepsilon_{it} \ . \tag{3.13}$$

obtained from (3.11a) by setting $I_t \stackrel{\text{def}}{=} \kappa - r^e_{\psi,t+1}$. Here, x_c and x_d denote the subvectors of x, comprised of its continuous and discrete components, respectively. The period-specific constant terms I_t are estimated by including appropriate time dummies. We choose λ_j for each of the J variables simultaneously by the following cross-validation criterion

$$\lambda^* = \arg\min_\lambda \sum_{t=1}^{T}\sum_{n=1}^{N_t}\left\{\Delta^0 p_{n,t} - \widehat{\Delta^0 p_{-n,t}}(\lambda)\right\}^2 , \tag{3.14}$$

where λ is the vector comprised of the λ_j for the different variables. Here, $\widehat{\Delta^0 p_{-n,t}}(\lambda)$ denotes the predicted value of $\Delta^0 p_{n,t}$ from an OLS fit of regression (3.13) using the transformations implied by λ but omitting the observation indexed (n,t) from the regression fit. By leaving out the observation used for evaluating the model fit the cross validated choice of λ^* is optimal in the sense of minimizing an estimate of the expected squared prediction error (Bunke et al.; 1999). Given the best transformations, we estimate (3.13) by OLS and use these estimates of I_t, β, σ_ε^2 to initialize the algorithm used for estimating the state space version of our model.

In general, a state space model (SSM) consists of a state and a measurement equation:

$$\alpha_t = T_t \alpha_{t-1} + \varepsilon_t^s \tag{3.15a}$$

$$y_t = Z_t \alpha_t + \varepsilon_t^m \,. \tag{3.15b}$$

with $\varepsilon_t^s \sim N(0, R_t)$ and $\varepsilon_t^m \sim N(0, H_t)$ (Harvey; 1989; Durbin and Koopman; 2001). The disturbance vectors are distributed independently. α_t denotes the unobservable vector of state variables, y_t the vector of observable measurements and T_t, Z_t, R_t and H_t are referred to as the system matrices. Our model (3.11) is easily written as a SSM. Let N_t denote the number of all houses sold at time t and let N denote the number of all observations. There are K_β house characteristics, and K_γ short run influence variables. $K = K_\beta + K_\gamma + 1$ is the number of constant state variables and $S = K + 1$ is the number of all state variables. We obtain

$$\alpha_t = \begin{bmatrix} r^e_{\psi,t+1} \\ \gamma \\ \kappa \\ \beta \end{bmatrix}, \; T_t = \left[\begin{array}{c:c:c} \phi & s_t & 0^\top_{K_\beta+1} \\ \hdashline 0_K & \multicolumn{2}{c}{I_K} \end{array}\right], \; \varepsilon_t^s = \begin{bmatrix} \xi_{\psi,t} \\ 0_K \end{bmatrix} \tag{3.16a}$$

$$y_t = \begin{bmatrix} \Delta^0 p_{1,t} \\ \vdots \\ \Delta^0 p_{N_t,t} \end{bmatrix}, \; Z_t = \begin{bmatrix} -1_{N_t} & 0_{N_t} 0^\top_{K_\gamma} & 1_{N_t} & X_t \end{bmatrix}, \; \varepsilon_t^m = \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{N_t,t} \end{bmatrix}, \tag{3.16b}$$

where I_K is a $(K \times K)$ identity matrix. Whereas the number of state variables per period is equal to S and fixed, the number of observations per period, i.e. N_t, varies. Note that the SSM primarily collects the unobservables of (3.11) in the state vectors α_t. Inferring α_t from the observed prices (stored in y_t) is the main econometric task. There are, however, three additional unknown parameters in other places of the SSM, namely ϕ in T_t, σ_ξ^2 in R_t and σ_ε^2 in H_t. These three parameters will be referred to as the unknown parameters of the SSM.

We are primarily interested in calculating the unobserved state vectors α_t. They contain the cumulated discount rate deviations $r^e_{\psi,t+1}$, the coefficients of the market indicators γ, and the influences of the characteristics β. If we knew all parameters of the SSM (3.15), we could use the Kalman smoother to figure out the state vectors. On the other hand if we knew α_t the parameters could be readily estimated by maximum likelihood given the distributional assumptions about ε^s_t and ε^m_t. The actual algorithm therefore iterates between the Kalman filter (to infer the α_ts) and the likelihood function (to estimate the unknown parameters of the SSM) until convergence is achieved. Appendix 3.6.1 presents the Kalman filter recursions.

Instead of using the log likelihood function of the SSM directly, we estimate the unknown parameters by maximizing the expected likelihood function using the EM algorithm with subsequent scoring steps (Dempster et al.; 1977; Watson and Engle; 1983; Shumway and Stoffer; 1982, 2000).

To set up the log-likelihood we multiply the system of the state equations with the S dimensional unit vector $e_1 = [1, 0, \dots, 0]^\top$. The log-likelihood is, up to a constant (Wu et al.; 1996)

$$\begin{aligned} \ln L(\psi) = & -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\varepsilon_0^\top \Sigma^{-1}\varepsilon_0 \\ & -\frac{1}{2}\sum_{t=1}^{T}\ln|\tilde{R}_t| - \frac{1}{2}\sum_{t=1}^{T}\tilde{\varepsilon}_t^{s\top}\tilde{R}_t^{-1}\tilde{\varepsilon}_t^s \\ & -\frac{1}{2}\sum_{t=1}^{T}\ln|H_t| - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t^{m\top}H_t^{-1}\varepsilon_t^m \end{aligned} \tag{3.17}$$

with $\varepsilon_0 = \alpha_0 - \mu$, $\tilde{R}_t \overset{\text{def}}{=} e_1^\top R_t e_1$, $\tilde{\varepsilon}^s_t = e_1^\top(\alpha_t - T_t\alpha_{t-1})$ and $\varepsilon^m_t = y_t - Z_t\alpha_t$. However, we do not observe the state vectors. The idea of the EM algorithm is to maximize instead the expected value of the log-likelihood function. To derive the expected value of (3.17), let us define for $t \leqslant T$

$$a_{t|T} \overset{\text{def}}{=} \mathcal{E}_T[\alpha_t] \tag{3.18a}$$

$$P_{t|T} \overset{\text{def}}{=} \mathcal{E}_T[(\alpha_t - a_{t|T})(\alpha_t - a_{t|T})^\top] \tag{3.18b}$$

$$P_{t,t-1|T} \overset{\text{def}}{=} \mathcal{E}_T[(\alpha_t - a_{t|T})(\alpha_{t-1} - a_{t-1|T})^\top]\,. \tag{3.18c}$$

Furthermore we rewrite

$$\varepsilon_0 = (\alpha_0 - a_{0|T}) + (a_{0|T} - \mu) \ ,$$

$$\tilde{\varepsilon}_t^s = e_1^\top \{(\alpha_t - a_{t|T}) - T_t(\alpha_{t-1} - a_{t-1|T}) + (a_{t|T} - T_t a_{t-1|T})\}$$

and

$$\varepsilon_t^m = (y_t - Z_t a_{t|T}) + Z_t(\alpha_t - a_{t|T}) \ .$$

We have for our model $H_t = \sigma_\varepsilon^2 I_{N_t}$ and $\tilde{R}_t = \sigma_\xi^2$. The assumption of uncorrelated errors in the discount rate and the price equation allows identification of the two variances (Schwann; 1998). After all, we obtain for (3.17) with $\mathcal{E}[\varepsilon^\top \Omega^{-1} \varepsilon] = \mathrm{tr}(\Omega^{-1}\mathcal{E}[\varepsilon\varepsilon^\top])$, where $\mathrm{tr}(A)$ denotes the trace of the square matrix A, i.e. the sum of A's diagonal elements,

$$\begin{aligned} \mathcal{E}_T[\ln L(\psi)] = & -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{tr}[\Sigma^{-1}\{P_{0|T} + (a_{0|T} - \mu)(a_{0|T} - \mu)^\top\}] \\ & - \frac{T}{2}\ln\sigma_\xi^2 - \frac{1}{2\sigma_\xi^2}\sum_{t=1}^{T} e_1^\top S_t e_1 - \frac{1}{2}\ln\sigma_\varepsilon^2 \sum_{t=1}^{T} N_t \\ & - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T}\mathrm{tr}(M_t) \end{aligned} \tag{3.19}$$

where

$$\begin{aligned} S_t \stackrel{\mathrm{def}}{=} \mathcal{E}_T[\varepsilon_t^s \varepsilon_t^{s\top}] = & \ P_{t|T} - P_{t,t-1|T}T_t^\top - T_t P_{t,t-1|T} + T_t P_{t-1|T} T_t^\top \\ & + (a_{t|T} - T_t a_{t-1|T})(a_{t|T} - T_t a_{t-1|T})^\top \end{aligned}$$

and

$$M_t \stackrel{\mathrm{def}}{=} \mathcal{E}_T[\varepsilon_t^m \varepsilon_t^{m\top}] = Z_t P_{t|T} Z_t^\top + (y_t - Z_t a_{t|T})(y_t - Z_t a_{t|T})^\top \ .$$

Due to the fact that the number of houses sold per period varies through time the filter procedure has to handle missing values. Generally, the Kalman filter is well suited for handling missing observations (Harvey; 1989, p. 144). One can either replace the missing observations with zeros and adjust the covariance matrix accordingly (Shumway and Stoffer; 2000, 4.4) or one can cancel out the missing observations from all matrices (Koopman et al.; 1999). It is possible to show that both methods deliver equivalent results, see Appendix 3.6.1. We use the second method in our XploRe algorithms `gkalarray`, `gkalfilter`, `gkalsmoother`, `gkallag` and `gkalresiduals`, see also the XploRe tutorial `gkalstart` and Schulz and Werwatz (2002).

The unknown parameters—collected in ψ—are μ, $\mathrm{vech}(\Sigma)$, ϕ, σ_ε^2 and σ_ξ^2, where $\mathrm{vech}(\Sigma)$ stags the lower triangular part of the square matrix Σ into a vector. We have to choose these parameters in such a manner that the value of the expected log-likelihood is maximized. It is easy to see that $\hat{\mu} = a_{0|T}$. We use the covariance matrix derived for the OLS estimates. For the other unknown coefficients we obtain with the help of the first order conditions

$$\hat{\sigma}_\xi^2 = \frac{1}{T}\sum_{t=1}^{T} e_1^\top S_t e_1$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{\sum_{t=1}^{T} N_t}\sum_{t=1}^{T} \mathrm{tr}(M_t)$$

$$\hat{\phi} = \frac{\sum_{t=1}^{T} e_1^\top \{P_{t,t-1|T} + a_{t|T}a_{t|T}^\top - T_{t,-\phi}(P_{t-1|T} + a_{t-1|T}a_{t-1|T}^\top)\}e_1}{\sum_{t=1}^{T} e_1^\top (P_{t-1|T} + a_{t-1|T}a_{t-1|T}^\top)e_1}\ , \tag{3.20}$$

where $T_{t,-\phi}$ is T_t, but ϕ is replaced by a zero. For the derivation of the expression for $\hat{\phi}$ we have to differentiate $e_1^\top S_t e_1$ with respect to ϕ, see (3.19). We use some results of vector and matrix differentiation (Lütkepohl; 1996, p. 208). We obtain for the relevant scalars with $dT_t/d\phi = dT_t^\top/d\phi = e_1 e_1^\top$

$$\frac{de_1^\top T_t P_{t,t-1|T} e_1}{d\phi} = e_1^\top P_{t,t-1|T} e_1\ ,$$

$$\frac{de_1^\top T_t P_{t-1|T} T_t^\top e_1}{d\phi} = 2e_1^\top T_t P_{t-1|T} e_1\ ,$$

$$\frac{de_1^\top T_t a_{t-1|T} a_{t|T}^\top e_1}{d\phi} = e_1^\top a_{t|T} a_{t-1|T}^\top e_1\ ,$$

$$\frac{de_1^\top T_t a_{t-1|T} a_{t-1|T}^\top T_t^T e_1}{d\phi} = 2e_1^\top T_t a_{t-1|T} a_{t-1|T}^\top e_1\ .$$

Thus, we obtain

$$\frac{de_1^\top S_t e_1}{d\phi} = 2e_1^\top (T_t P_{t-1|T} + T_t a_{t-1|T} a_{t-1|T}^\top - P_{t,t-1|T} - a_{t|T} a_{t-1|T}^\top) e_1\ . \tag{3.21}$$

Eventually, we rewrite the half of the right-hand-side of (3.21) with

$$\begin{aligned} e_1^\top T_t &= e_1^\top(\phi e_1 e_1^\top + T_{t,-\phi}) \\ &= \phi e_1^\top + e_1^\top T_{t,-\phi} \end{aligned}$$

as

$$\begin{aligned} \phi e_1^\top(P_{t-1|T} + a_{t-1|T}a_{t-1|T}^\top)e_1 + e_1^\top\{T_{t,-\phi}(P_{t-1|T} + a_{t-1|T}a_{t-1|T}^\top) \\ - P_{t,t-1|T} - a_{t|T}a_{t-1|T}^\top\}e_1 \end{aligned}$$

and use this expression for the derivation of the third equation in (3.20). Moreover, one can derive that the second-order cross partial derivatives of the expected log likelihood function are zero at the stationary point $(\hat{\sigma}_\xi^2, \hat{\sigma}_\varepsilon^2, \hat{\phi})$. One obtains with

$$\frac{d^2 e_1^T S_t e_1}{d\phi^2} = 2e_1^\top(P_{t-1|T} + a_{t-1|T}a_{t-1|T}^\top)e_1 \geqslant 0$$

that the own partial derivatives are all negative. Thus, the values $(\hat{\sigma}_\xi^2, \hat{\sigma}_\varepsilon^2, \hat{\phi})$ also fulfill the second order condition for a local maximum.

The EM algorithm consists of the following iterative procedure: start with some reasonable values for the unknown coefficients, evaluate the matrices in the expected log-likelihood function with the Kalman smoother, and estimate the unknown coefficients. Use these estimates for a new evaluation of the expected log-likelihood and so on until convergence is reached. As Harvey (1989, p. 126) shows, it is possible to rewrite the log-likelihood (3.17) function in the prediction error decomposition form

$$\ln L(\psi) = -\frac{1}{2}\sum_{t=1}^{T} \ln|F_t| - \frac{1}{2}\sum_{t=1}^{T} v_t^\top F_t^{-1} v_t \tag{3.22}$$

with $v_t \overset{\text{def}}{=} y_t - Z_t a_{t|t-1}$. The matrix F_t is a by-product of the Kalman filter. In the above log-likelihood function we have omitted the expression for $t = 0$ and once again the constant term. The latter is $-0.5N\ln(2\pi)$ and depends solely on the total number of observations. The EM algorithm guarantees that the value of the likelihood increases for every iteration. However, it is a drawback of the algorithm that it does not deliver an estimate of the information matrix. This matrix is necessary to calculate standard errors for the estimated coefficients. Thus, we complete the estimation procedure with a final scoring step for (3.22) evaluated at the estimates of the EM algorithm. As Engle and Watson (1981) have shown, the elements of the information matrices are given by (with $i, j = 1, \ldots 3$)

$$\mathcal{I}_{ij} = \sum_{t=1}^{T} \left\{ \frac{1}{2} \mathrm{tr} \left(F_t^{-1} \frac{\partial F_t}{\partial \psi_i} F_t^{-1} \frac{\partial F_t}{\partial \psi_j} \right) + \left(\frac{\partial v_t}{\partial \psi_i} \right)^{\top} F_t^{-1} \frac{\partial v_t}{\partial \psi_j} \right\} .$$

which can be evaluated numerically. All computational routines are implemented as Quantlets in XploRe (Härdle et al.; 2000) and use the generalized Kalman filter routines form its `kalman` library.

3.4 Empirical investigation

3.4.1 Data

The primary data set is provided by Berlin's Surveyor Commission for Real Estate (GAA). According to the German Building Law (BauGB) notaries are obliged to sent copies of contracts for sale of properties to the Surveyor Commission in their respective state. Surveyor Commissions have to store the data and use it to provide information on the real estate market (§§ 192-199 BauGB). Our data set is stored in the non-public sector of MD*Base, www.mdtech.de, and contains information on 4410 sales of single-family houses that occurred between August 1982 and December 1999 in the four South-West districts of Berlin. Since we chose the month as the time period of our analysis, this gives us at least 6, at most 43 and on average 21 observations per period. Each observation consists, in addition to the price, of about 60 variables such as the size of the lot, floor space, age of the house, location, availability and numerous qualitative variables indicating specific conditions of the house, the neighborhood or the transaction (e.g. transaction between relatives). Our primary data set is the source for most of the variables contained in the price equation (3.11a).

We collected additional information about tax rates and government housing programs during the relevant time period to construct appropriate dummy variables representing rules changes that may influence potential home owners decisions to buy a house.

The four South-West districts cover 19% of Berlin's area, see the map given in Figure 3.1 where the four South-West districts are shaded and accounted in 1998 for 17% of Berlin's total population of about 3.4 million. The

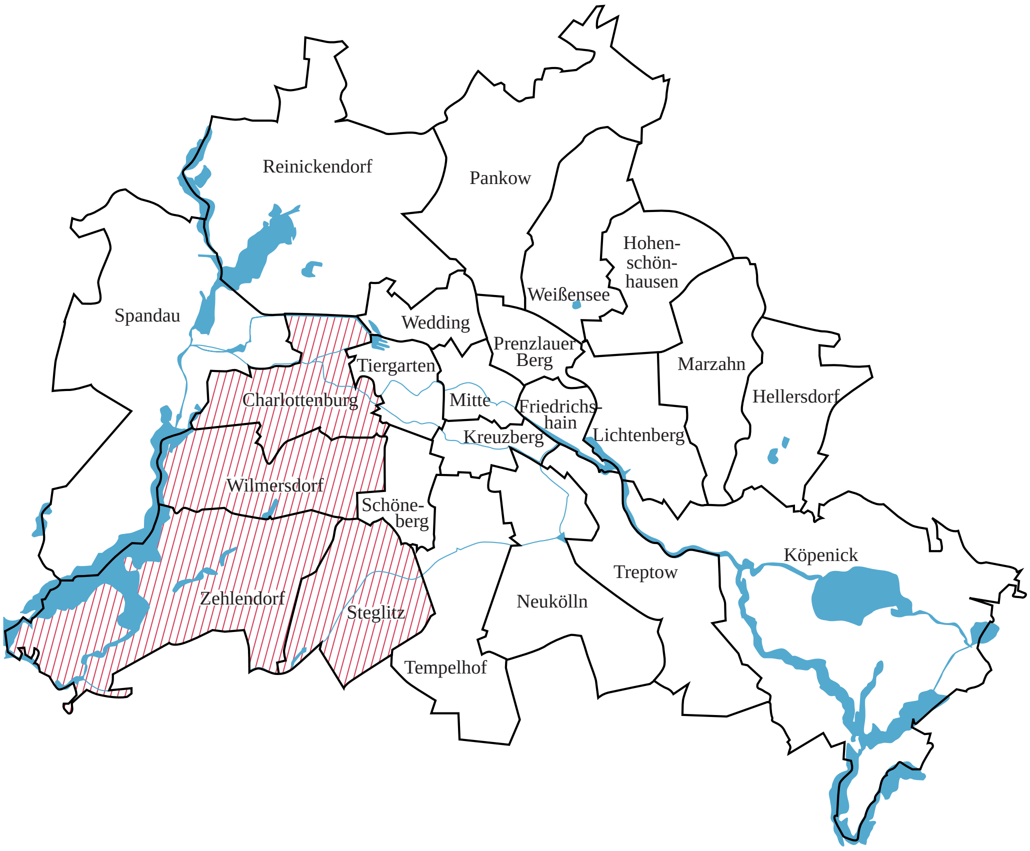

Figure 3.1: Berlin's districts. Our model is fitted for sales from the four shaded South-West districts.

districts are of high-quality and relatively homogeneous. It is reasonable that houses in these districts share the same market risk, so that imputed rents of house ownership will be discounted by the same rate. Moreover, these districts are part of the former West Berlin and thus might be influenced to a lesser degree by the effects of the reunification of the city and the country in 1990, which produced a major boom in Berlin's real estate market. Table 3.1 reports summary statistics for the transacted houses.

The dependent variable $\Delta^0 p_{n,t}$ of the price equation is the deviation of the sale price $p_{n,t}$ from the value of the reference dwelling at time t, $\hat{p}_t^0$. Since $\hat{p}_t^0$ is not directly observable, we estimate it according to the steps described in Appendix 3.6.2, using the monthly rent sub-aggregate of the consumer

Table 3.1: *Summary statistics for transacted single-family houses in the four South-West districts of Berlin between 1982:08 to 1999:12.*

Panel A: Continuous Characteristics and Prices						
	Mean	Median	Std. Dev.	Min	Max	Units
Lot size	600.9	534.0	361.5	127.0	6670.0	*Square metres*
Floor space	169.5	147.0	72.0	45.0	754.0	*Square metres*
Age	41.9	46	21.9	0	129	*Years*
Price	387.1	306.8	295.3	38.3	4192.6	*Thsd. EUR*
Panel B: Availability at the Date of Sale						
Under construction			1.1%	Rented out		1.3%
Purchase by former tenant			2.7%	Seller-occupied		94.9%
Panel C: House Types and Districts						
Detached			48.5%	Charlottenburg		5.3%
Row			27.5%	Steglitz		37.4%
Semi-detached			24.0%	Wilmersdorf		7.0%
Observations			4410	Zehlendorf		50.3%

Note: Original currency units are German marks which are converted to EUR by dividing with 1.95583.

price index for Berlin. The plot of the estimated series in Figure 3.2 reveals that the reference value soars both in the early Eighties as well as in the first half of the Nineties, reaching its peak in 1995. The value remains on a relatively constant level thereafter. The sharp increase of estimated price of the reference dwelling in the early Nineties reflects the boom in Berlin's real estate market following the reunification of Germany and the concurrent decision to move the nation's capital from Bonn to Berlin. The mean of the dependent variable $\Delta^0 p_{n,t}$ is 8.53 and its standard deviation is 0.50.

Regarding the return deviation equation (3.11b), we use different market indicators in the fashion of Chinloy et al. (1997); Ling and Naranjo (1997) as components of the vector s_t. Appendix 3.6.3 describes the data in detail. Positive inflation shocks should decrease the required discount rate, because properties are real goods which provide a hedge against adverse shocks, see the results of Engle et al. (1985) and the results from Section 2.2. The future composition of the market for dwellings might have an influence on the discount rate. We measure the future composition by the ratio of building permissions for single-family houses to the total number of building permissions for residential buildings in Berlin West. An above average ratio should have a positive effect on the discount rate. Spreads between mortgage rates and interest rates with the same maturity are measures of the risk premiums that banks command for home loans (Nautz and Wolters; 1996). Higher spreads indicate higher risk perception and should increase the discount rate. Eventually, returns on alternative investments may have an influence on the discount rate. The returns of the CDAX index are used to capture the performance of the German stock market.

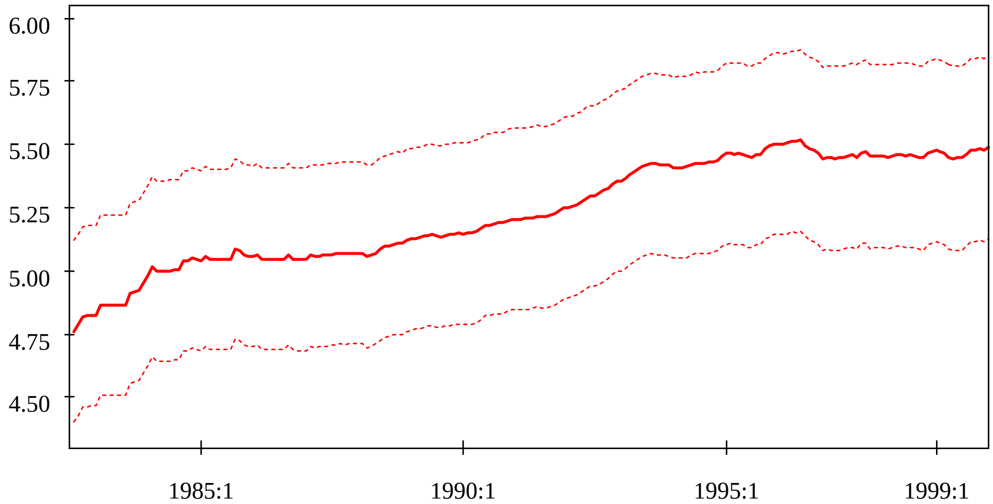

Figure 3.2: Plot of the reference value $\hat{p}_t^0$ from 1982:8 to 1999:12, which is defined in equation (3.6). Confidence intervals are calculated with the delta method, see Appendix 3.6.2.

3.4.2 Empirical results

As described in Section 3.3 our empirical analysis basically proceeds in two steps. In the first step, we select the functional form of (3.11a) in a one-equation hedonic regression approach. In the final step, we use the OLS estimates of the optimal one-equation fit as starting values for estimating all unknowns of the two-equation system (3.11) in the state space framework.

Model selection and OLS estimates

Working with the standard hedonic price equation (3.13) we determined the optimal transformations of the three continuous components of $x_{n,t}$ (size of the lot, size of the floor space and age of the building) and chose the significant discrete explanatory variables from a larger set by a backward selection procedure. The optimal values of λ are 1 (size of the lot), -2 (size of the floor space), and -0.5 (age of the building), respectively. The value of the R^2-like, standardized cross-validation criterion in (3.14) for these transformations is 0.758.

Among the variables eliminated from the model were the indicators of general changes in the tax- and subsidy rules. Generally, it is not clear to which extend subsidies and taxes will be capitalized into house prices (Bruce and Holtz-Eakin; 1999; Hendershott and White; 2000). Impacts will depend directly on the situation of buyer and seller and are only identifiable with information on the latter. Berger et al. (2000) have found evidence for Swedish data that interest subsidies are capitalized into transaction prices. However, these subsidies are tied to the house (rather than to the buyer or the seller). German subsidies for owner-occupied housing are means tested and increase with the number of kids living in the household. Couples can apply for such subsidies two times in their life. Our results, however, suggest at least that general changes in German tax- and subsidy rules had no influence on transaction prices. Major changes are the repealing of taxation of imputed rents in 1987 and the change in promotion of owner-occupied housing by direct subsidies in 1996. It was the intention of both changes to continue pre-existing rules, which is confirmed by insignificant dummy coefficients for all owner-occupied houses sold before 1987 and after 1995, respectively.

The final model thus consists of the three transformed continuous and the sixteen discrete characteristics whose estimated coefficients are reported in Table 3.13. For the following interpretation, we should recall that an increase in a characteristic of a house changes the imputed rent by the change weighted with the respective implicit price, see equation (3.4). The price, as the sum of the discounted imputed rents, changes by the same amount, see equation (3.9). However, all characteristics that describe circumstances of the sale influence only the price. We find that the price for a house increases both with the size of the lot and the size of the living area and decreases with the age of the dwelling. Evaluated at the respective sample means given in Table 3.1, we obtain elasticities of about 0.3% for lot size, 0.6% for floor

Table 3.2: *OLS estimates of optimal specification of the hedonic equation for single-family houses 1982:8 to 1999:12.*

	Coefficient	t-Statistic	P-Value
T_1(lot size)	0.178	32.84	0.000
T_{-2}(floor space)	26.224	44.08	0.000
$T_{-0.5}$(age)	-0.042	-12.42	0.000
Row house	-0.035	-3.33	0.001
Detached house	0.076	7.21	0.000
Wilmersdorf	0.247	11.42	0.000
Zehlendorf	0.098	5.65	0.000
Steglitz	-0.105	-5.89	0.000
Good condition	0.116	13.74	0.000
Bad condition	-0.230	-6.66	0.000
Noise	-0.268	-3.27	0.001
Indoor pool	0.097	3.34	0.001
Fixtures	0.080	4.00	0.000
Under construction	-0.221	-4.69	0.000
Rented out	-0.153	-4.60	0.000
Purchased by former tenant	-0.112	-4.57	0.000
Personal circumstances	-0.163	-9.04	0.000
Unusual circumstances	-0.157	-5.29	0.000
	Diagnostics		
R^2	0.783	$\bar{R}^2$	0.771
F-Statistic	66.706	P-Value(F-Stat.)	0.000
Observations	4410	$\widehat{\sigma_\varepsilon^2}$	0.057

Notes: Dependent variable is the log ratio of price to reference value. $T_\lambda(\cdot)$ is the transformation function given in equation (3.12). Included overall constant and time dummies are not reported.

space, and -0.03% for age. These figures are comparable with numbers from other studies.

The price drops by 3.5% for row houses and rises by 7.6% for detached houses compared with a semi-detached house. As such, people are willing to pay a premium for 'privacy'. They will also pay sizeable premiums of 24.7% and 9.8% if the house lies in the districts Wilmersdorf or Zehlendorf, respectively. Wilmersdorf and Zehlendorf have very nice sections with forests and lakes, see Figure 3.1, and contain Berlin's two finest neighborhoods Grunewald and Wannsee.

The price rises by 11.6% for houses in good condition and declines by 23.0% for houses in bad condition compared with a house in normal condition. If the house is located in a noisy environment in the vicinity of rail tracks, highways, or airports, its price decreases by 26.8%. The price increases by 9.7% if the object has an indoor pool and by 8.0% if it has fixtures such as built-in kitchen, built-in furniture or a sauna. The average price for houses with an indoor pool is 562,421 EUR and the text files of our data reveal that constructing costs for an indoor pool go up to 51,100 EUR.

Next we turn to the dummies that describe special circumstances of the deal or the use of the house that are only relevant for house buyers. If the house is still under construction when the deal is struck, the 'risk' rebate for buying an unfinished house is about 22.1% of the price for an otherwise identical object. When the buyer is an investor who wants to accrue payments from a rented house, the rebate of 15.3% captures higher operating cost. Expert-rated additional cost for rented properties are 2% for default risk, 5% for administration cost, so that maintenance cost are 8.3% higher due to the principal agent problem between lessor and lessee. It is also possible that the buyer would like to move into the house. While this, in principle, suffices to terminate the lease according to the German law, it may only occur with a considerable delay or may not occur at all if the tenant reclaims undue hardship. If the house is sold to the former tenant, the rebate amounts to 11.2% only because the tenant has a special interest in the house—it is arranged according to his taste—and because he saves moving costs. If the transaction is influenced by personal circumstances, the price decreases by 16.3%. This category comprises sales between relatives, where bequest motifs might explain the rebate, and sales between divorced couples and partitioning of an estate, where there might not be enough time and patience for getting a good deal. In addition, it also contains sales between neighbors, where the seller incurs no search and information cost. Finally, unusual circumstances with respect to the business dealing command a rebate of about 15.7%. This is reasonable for sales where the former owner has obtained the right of residence in parts of the house and when the buyer has to repay in kind such as providing nursing care for the former owner.

Estimates of the state space model

Using the specification of the price equation selected in the preliminary hedonic regression analysis, we estimate the SSM by combining the Kalman filter with the EM algorithm. After achieving convergence, we perform additional scoring steps based on the prediction error decomposition form of the log-likelihood function (Harvey; 1989) to get asymptotically efficient estimates of the parameters in the SSM. The output consists of the smoothed state vectors α_t and the parameter estimates of ϕ, $\ln \sigma_\xi$ and $\ln \sigma_\varepsilon$, where the log transformations ensure numerical stability and positive estimates. Regarding the state vectors, note from (3.16a) that only the first component is time varying, whereas the components γ, κ and β are time constant. The values of the time constant components are reported in Table 3.3 except for the estimates of the implicit prices β because they are only slightly different

Table 3.3: *Estimates of the parameters in the SSM and influences of the market indicators.*

	Coefficient		t-Statistic	P-Value
$\hat{\phi}$	0.925		29.23	0.000
$\widehat{\ln \sigma_\xi}$	-3.757		-26.19	0.000
$\widehat{\ln \sigma_\varepsilon}$	-1.437		-132.05	0.000
$\hat{\kappa}$	-4.412		-214.36	0.000
Unexpected inflation	-0.889		-0.81	0.419
Ratio of building permissions	0.030		4.78	0.000
Spread10	6.308		3.28	0.001
CDAX return	0.025		0.48	0.633
Diagnostics				
Log likelihood	4082.612	$\widehat{\sigma_\varepsilon^2}$		0.056
Observations	4410	$\widehat{\sigma_{\Delta p}^2}$		0.250

Note: The market indicators are lagged by one month and demeaned. Appendix 3.6.3 describes the indicators.

from the OLS estimates given in Table 3.2. Table 3.3 also contains the MLE estimates of the parameters of the SSM. The distribution of the standardized residuals is unimodal, and resembles the normal distribution, but possess—as a QQ plot reveals—heavier tails. The variance of the disturbances accounts for 22% of the total variation of the log ratios.

As in Engle et al. (1985) the estimate of ϕ implies that the common price component reacts sluggish and may have a unit root. This, in general, does not pose a problem for our estimation strategy (Engle and Watson; 1981). However, an instationary common price component is at odds with our interpretation as deviations from the long run discount rate. The effects of the German Reunification might have temporarily suspended the equilibrating process between the reference value—and thus the rents—and prices for single-family houses. Taking an agnostic view, we still interpret our results under the assumption of stationary return deviations.

The estimated series of the expected return deviations $r^e_{\psi,t}$ is plotted in Figure 3.3, along with 95% confidence intervals which suggest that the return deviation was zero for the first years of our sample period. In 1985, the discount rate started to increase and the price for the reference dwelling was lower than the corresponding reference value. This down-weighting process reached its peak in 1987. Thereafter, the prices—compared with the reference value—increased steadily. Starting in 1990, investor's confidence reached very high levels and prices increased substantially. There are at least three complementary explanations for this surge in confidence: the Economic, Currency and Social Union in July 1990, the German Reunification in October 1990, and the decision of the German Parliament in June 1991 to make Berlin the Capital of the unified country. All these events—or their anticipation—may have contributed as shocks to the decrease in the required returns. To evaluate if the return deviations are of a sensible magnitude, we used our implied estimate of ψ, which is about 13. Recall that according to the Gordon growth model $r^* = \ln(1+G) + \ln(1+\theta)$. We obtain $r^* = 0.815\%$ by using $G = 0.4\%$, see Table 3.4, and the median monthly θ of 0.415%. To guarantee plausible return deviations, we should have $r^e_t + r^* > 0$ and thus $r^e_{\psi,t} > -0.1$. Even if we use the upper limit of the confidence bands, the return deviations are below that critical value from February 1991 to February 1993. The lowest upper bound is about -0.2 and thus r^* should be 1.5% to guarantee that the discount rate is always positive. In this 'worst' case, the implied cap rate θ will be 1.1%, a value that is still in the range found by Engle et al. (1985) in their study on single-family house prices in a suburb of San Diego. Comparing with the behavior of the

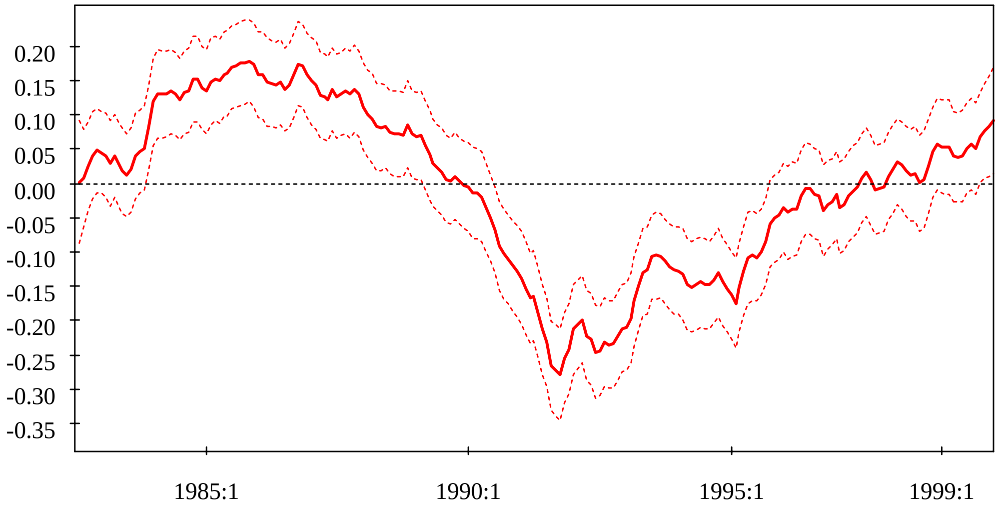

Figure 3.3: Plot of the estimated expected return deviations, $r^e_{\psi,t+1}$, from 1982:8 to 1999:12. Confidence intervals are calculated with the first element of the smoothed covariance matrix.

reference value in Figure 3.2, we see that the rents reacted apparently more slowly to this new situation. By 1996 though, the series of return deviations was back to zero, implying that the current return had returned once again to its long-run equilibrium level.

One of the strength of the state spate formulation of our two-equation model (3.11) is that it allows to estimate the effects of observable market indicators on the common price component. The signs of the estimated coefficients of the market indicators, given in Table 3.3, are reasonable in sign. Unexpected inflation decreases the required return rate. With positive inflation shocks, the attractiveness of investments in real estate increases. The estimated coefficient, however, is not statistically significant at the 5% level. If the relative supply of new single-family houses increases with respect to the supply of residential buildings—measured by the log ratio of building permissions—the required returns increase. Higher risk premiums for investments in real estate—measured by the spread between mortgage and interest rates—also have a positive influence on the discount rate, while the return of the stock market index CDAX has no significant influence on the short run discount rate.

3.5 Conclusion

In this chapter, we have estimated an empirical model of house prices that extends the conventional hedonic regression approach in several important respects. Our model includes a regression equation relating house prices to observable characteristics and an unobservable price component, much like in a standard hedonic model. However, we derive and interpret this equation using the present value model of asset prices and carefully select its functional form by combining stepwise regression and cross validation techniques in the empirical analysis. We augment this price equation by an autoregressive model for the unobserved component of prices that is common to all houses sold in this period. The theoretical framework suggests to interpret this component as the deviation of the return rate for investments in single-family homes from its long-run equilibrium level.

The empirical specification allows the law of motion of this quantity to depend both on observable and unobservable factors. It is estimated by putting the two-equation model into state space form and applying the Kalman filter to infer the common price component from the observable house prices. The Kalman filter is combined with maximum likelihood estimation based on the EM algorithm to obtain estimates of all unknown parameters of our model.

We carefully interpret the coefficients of the price equation and find the estimated hedonic coefficients to be both plausible in sign and magnitude and to be in accordance with the assessment of professional appraisers. The estimated series of the deviations of the return rate of housing investment suggests that investors were overconfident following the German reunification and the subsequent decision to make Berlin the Capital of the unified Germany. While our model, based on a constant long-run discount rate, has problems to capture this highly speculative period it nonetheless illustrates the usefulness of modelling the unobserved price equation with a separate equation.

3.6 Appendix

3.6.1 Kalman filter recursions

Here we explain in detail the calculation procedure of the Kalman filter and the Kalman smoother. For a derivation of the recursions we refer to Harvey (1989).

Calculation procedure for the Kalman filter

Start at $t = 1$: using an initial guess for μ and Σ to calculate

$$a_{1|0} = T_1 \mu \ , \quad P_{1|0} = T_1 \Sigma T_1^\top + R_1 \ , \quad F_1 = Z_1 P_{1|0} Z_1^\top + H_1$$

$$a_1 = a_{1|0} + P_{1|0} Z_1^\top F_1^{-1} (y_1 - Z_1 a_{1|0})$$
$$P_1 = P_{1|0} - P_{1|0} Z_1^\top F_1^{-1} Z_1 P_{1|0}$$

Step at $t \leqslant T$: calculate with a_{t-1} and P_{t-1}

$$a_{t|t-1} = T_t a_{t-1}$$

$$P_{t|t-1} = T_t P_{t-1} T_t^\top + R_t \ , \quad F_t = Z_t P_{t|t-1} Z_t^\top + H_t$$
$$a_t = a_{t|t-1} + P_{t|t-1} Z_t^\top F_t^{-1} (y_t - Z_t a_{t|t-1})$$
$$P_t = P_{t|t-1} - P_{t|t-1} Z_t^\top F_t^{-1} Z_t P_{t|t-1}$$

The recursions are implemented in XploRe in the Quantlet `gkalfilter`. In addition to the above recursions, it also calculates the log likelihood function given in equation 3.22.

Calculation procedure for the Kalman smoother

To run the Kalman smoother, one needs a_t, P_t and $P_{t|t-1}$ for $t = 1 \ldots T$ from the previous procedure.

Start at $t = T$:

$$a_{T|T} = a_T$$
$$P_{T|T} = P_T$$

Step at $t \leqslant T-1$: calculate with $a_{t+1|T}$ and $P_{t+1|T}$

$$P_t^* = P_t T_{t+1}^\top P_{t+1|t}^{-1}$$

$$a_{t|T} = a_t + P_t^*(a_{t+1|T} - T_{t+1}a_t)$$
$$P_{t|T} = P_t + P_t^*(P_{t+1|T} - P_{t+1|t})P_t^{*\top}$$

The recursions are implemented in XploRe in the Quantlet `gkalsmoother`.

We need furthermore a smoothed series for $P_{t,t-1|T}$. The recursions are, see Shumway and Stoffer (1982, 2000),

Start at $t = T$:

$$P_{T,T-1|T} = \{I_S - P_{T|T-1}Z_T^\top(Z_T P_{T|T-1} Z_T^\top + H_T)^{-1} Z_T\} T_T P_{T-1}$$

Step at $t < T-1$: calculate

$$P_{t,t-1|T} = \{P_t + P_t^*(P_{t+1,t|T} - T_{t+1}P_t)\}P_{t-1}^{*\top}$$

Here, I_S is a $(S \times S)$ identity matrix. The recursions are implemented in XploRe in the Quantlet `gkallag`.

Procedure equivalence

We show that our treatment of missing values delivers the same results as the procedure proposed by Shumway and Stoffer (1982, 2000). For this task, let us assume that the $(N \times 1)$ vector of observations t

$$y_t^\top = \begin{bmatrix} y_{1,t} & . & y_{3,t} & . & y_{5,t} & \dots & y_{N,t} \end{bmatrix}$$

has missing values. Here, observations 2 and 4 are missing. Thus, we have only $N_t < N$ observations. For Kalman filtering in XploRe, all missing values in y_t and the corresponding rows and columns in the measurement matrices d_t, Z_t, and H_t, are deleted. Thus, the adjusted vector of observations is

$$y_{t,1} = \begin{bmatrix} y_{1,t} & y_{3,t} & y_{5,t} & \dots & y_{N,t} \end{bmatrix}$$

where the subscript 1 indicates that this is the vector of observations used in the XploRe routines. The procedure of Shumway and Stoffer instead rearranges the vectors in such a way that the first N_t entries are the observations—and thus given by $y_{t,1}$—and the last $(N - N_t)$ entries are the missing values. However, all missing values must be replaced with zeros.

For our proof, we use the following generalized formulation of the measurement equation

$$\begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} = \begin{bmatrix} d_{t,1} \\ d_{t,2} \end{bmatrix} + \begin{bmatrix} Z_{t,1} \\ Z_{t,2} \end{bmatrix} \alpha_t + \begin{bmatrix} \varepsilon^m_{t,1} \\ \varepsilon^m_{t,2} \end{bmatrix}$$

and

$$\operatorname{cov}\begin{pmatrix} \varepsilon^m_{t,1} \\ \varepsilon^m_{t,2} \end{pmatrix} = \begin{bmatrix} H_{t,11} & H_{t,12} \\ H_{t,12} & H_{t,22} \end{bmatrix} .$$

$y_{t,1}$ contains the observations and $y_{t,2}$ the missing values. The procedure of Shumway and Stoffer uses the above given generalized formulation and sets $y_{t,2} = 0$, $d_{t,2} = 0$, $Z_{t,2} = 0$, and $H_{t,12} = 0$ (Shumway and Stoffer; 2000, p. 330). We should remark that also the dimensions of these matrices depend on t via $(N - N_t)$. However, to guarantee a simple notation we omit any indexation of this time dependency. It is important to mention that matrices with subscript 1 and 11 are equivalent to the adjusted matrices of XploRe's filtering routines.

First, we show that both procedures deliver the same results for the Kalman filter. We do this by induction. If this is shown, we conclude that the smoother also delivers identical results.

Proof. Given μ and Σ, the terms $a_{1|0}$ and $P_{1|0}$ are the same for both procedures. This follows from the simple fact that the first two steps of the Kalman filter do not depend on the vector of observations, see the above given recursions of the Kalman filter.

Now, given $a_{t|t-1}$ and $P_{t|t-1}$, we have to show that also the filter recursions

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t^\top F_t^{-1} v_t \ , \quad P_t = P_{t|t-1} - P_{t|t-1} Z_t^\top F_t^{-1} Z_t P_{t|t-1} \tag{3.23}$$

deliver the same results. Let ss label the results of the Shumway and Stoffer procedure. We derive with

$$Z_{t,ss} \stackrel{\text{def}}{=} \begin{bmatrix} Z_{t,1} \\ 0 \end{bmatrix}$$

that

$$F_{t,ss} = \begin{bmatrix} Z_{t,1} P_{t|t-1} Z_{t,1}^\top & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} H_{t,11} & 0 \\ 0 & H_{t,22} \end{bmatrix} .$$

The inverse is (Sydsæter et al.; 2000, 19.49)

$$F_{t,ss}^{-1} = \begin{bmatrix} F_{t,1}^{-1} & 0 \\ 0 & H_{t,22}^{-1} \end{bmatrix} \tag{3.24}$$

where $F_{t,1}$ is just the covariance matrix for the innovations of XploRe's procedure. With (3.24) we obtain that

$$Z_{t,ss}^{\top} F_{t,ss}^{-1} = \begin{bmatrix} Z_{t,1}^{\top} F_{t,1}^{-1} & 0 \end{bmatrix}$$

and accordingly for the innovations

$$v_{t,ss} = \begin{bmatrix} v_{t,1} \\ 0 \end{bmatrix} .$$

We obtain immediately

$$Z_{t,ss}^{\top} F_{t,ss}^{-1} v_{t,ss} = Z_{t,1}^{\top} F_{t,1}^{-1} v_{t,1} .$$

Plugging this expression into (3.23)—taking into account that $a_{t|t-1}$ and $P_{t|t-1}$ are identical—delivers

$$a_{t,ss} = a_{t,1} \quad \text{and} \quad P_{t,ss} = P_{t,1} .$$

This completes the first part of our proof.

The Kalman smoother recursions use only system matrices that are the same for both procedures. In addition to the system matrices, the output of the filter is used as input. But we have already shown that the filter output is identical. Thus the results of the smoother are the same for both procedures as well. □

Smoothed constant state variables

We want to show that the Kalman smoother produces constant estimates through time for all state variables that are constant by definition. To proof that, we use some of the above given smoother recursions.

First of all, we rearrange the state vector such that the last $s \leqslant S$ variables are constant. This allows the following partition of the transition matrix

$$T_{t+1} = \begin{bmatrix} T_{11,t+1} & T_{12,t+1} \\ 0 & I_s \end{bmatrix} \tag{3.25}$$

with the $s \times s$ identity matrix I_s. Furthermore, we define with the same partition

$$\tilde{P}_t \stackrel{\text{def}}{=} T_{t+1} P_t T_{t+1}^\top = \begin{bmatrix} \tilde{P}_{11,t} & \tilde{P}_{12,t} \\ \tilde{P}_{12,t} & \tilde{P}_{22,t} \end{bmatrix}$$

The filter recursion for the covariance matrix are given as

$$P_{t+1|t} = T_{t+1} P_t T_{t+1}^\top + R_{t+1}$$

where the upper left part of R_{t+1} contains the covariance matrix of the disturbances for the stochastic state variables. We see immediately that only the upper left part of $P_{t+1|T}$ is different from $\tilde{P}_t$.

Our goal is to show that for the recursions of the smoother holds

$$P_t^* = \begin{bmatrix} M_{11,t} & M_{12,t} \\ 0 & I_s \end{bmatrix} , \tag{3.26}$$

where both Ms stand for some complicated matrices. With this result at hand, we obtain immediately

$$a_{t|T}^s = a_{t+1|T}^s = a_T^s \tag{3.27}$$

for all t, where $a_{t|T}^s$ contains the last s elements of the smoothed state $a_{t|T}$.

Furthermore, it is possible to show with the same result that the lower right partition of $P_{t|T}$ is equal to the lower right partition of P_T for all t. This lower right partition is just the covariance matrix of $a_{t|T}^s$. Just write the smoother recursion

$$P_{t|T} = P_t (I_S - T_{t+1}^\top P_t^{*\top}) + P_t^* P_{t+1|T} P_t^{*\top} .$$

Then check with (3.25) and (3.26) that the lower-right partition of the first matrix on the right hand side is a $s \times s$ matrix of zeros. The lower-right partition of the second matrix is given by the the lower-right partition of $P_{t+1|T}$.

Proof. Now we derive (3.26): We assume that the inverse of T_{t+1} and $T_{11,t+1}$ exist. For the partitioned transition matrix (Sydsæter et al.; 2000, 19.48) we derive

$$T_{t+1}^{-1} = \begin{bmatrix} T_{11,t+1}^{-1} & -T_{11,t+1}^{-1} T_{12,t+1} \\ 0 & I_s \end{bmatrix} . \tag{3.28}$$

Now, it is easy to see that

$$P_t^* = T_{t+1}^{-1} \tilde{P}_t P_{t+1|t}^{-1} \; . \tag{3.29}$$

We have (Sydsæter et al.; 2000, 19.49)

$$P_{t+1|t}^{-1} = \begin{bmatrix} \Delta_t & -\Delta_t \tilde{P}_{12,t} \tilde{P}_{22,t}^{-1} \\ -\tilde{P}_{22,t}^{-1} \tilde{P}_{12,t} \Delta_t & \tilde{P}_{22,t}^{-1} + \tilde{P}_{22,t}^{-1} \tilde{P}_{12,t} \Delta_t \tilde{P}_{12,t} \tilde{P}_{22,t}^{-1} \end{bmatrix} \tag{3.30}$$

with Δ_t as a known function of the partial matrices. If we multiply this matrix with the lower partition of $\tilde{P}_t$ we obtain immediately $[0 \; I_s]$. With this result and (3.28) we derive (3.26). □

3.6.2 Reference value

For calculating the reference value we use the monthly rent sub-aggregate of the consumer price index for Berlin, provided by Berlin's Statistical Office (StaLa) in its Statistical Report M I 2. We fit the following regression to the series of log differences

$$\Delta d_t^0 = \delta_0 + \delta_1 \Delta d_{t-1}^0 + \delta_2 \Delta d_{t-12}^0 + u_t^0 \; . \tag{3.31}$$

Using the Ljung-Box Q-statistic up to 36 lags, we cannot reject the null that the residuals are uncorrelated with a probability value of about 0.38. Table 3.4 reports the results. Rewriting (3.31) as a VAR(1) gives

$$v_t = c + A v_{t-1} + u_t \tag{3.32}$$

where the 13×1 vector v_t contains the observations of Δd_t^0 from t to $t-12$. For c we have $c_1 = \delta_0$ and all other elements are zero. Furthermore, $a_{1,1} = \delta_1$, $a_{1,13} = \delta_2$, $a_{j,j-1} = 1$ for $j = \{2, .., 13\}$ and all other elements are zero. Finally, the first element in u_t is the noise term u_t^0 and all other elements are zero. The matrix A has 13 distinct eigenvalues which all have modulus less than 1. Using the unit vector $e_1 = [1, 0, \ldots, 0]^\top$, we obtain

$$\mathcal{E}_t\left[\Delta d_{t+1+j}^0\right] = e_1^\top \left(\sum_{i=0}^{j} A^i \right) c + e_1^\top A^{j+1} v_t \quad \text{for} \quad j \geqslant 0 \; .$$

According to (3.6), these changes have to be discounted by ρ, which gives

$$\sum_{j=0}^{\infty} \rho^j \mathcal{E}_t\left[\Delta d_{t+1+j}^0\right] = \frac{1}{1-\rho} e_1^\top (I - \rho A)^{-1} c + e_1^\top A (I - \rho A)^{-1} v_t \; . \tag{3.33}$$

Table 3.4: *Regression results for the process of the rent index from 1980:2 to 1999:12.*

	Coefficient		t-Statistic	P-Value
$\widehat{\delta_0}$	0.001		3.09	0.002
$\widehat{\delta_1}$	0.145		2.60	0.009
$\widehat{\delta_2}$	0.506		8.98	0.000
		Diagnostics		
R^2	0.301	mean of Δd^0		0.004
$\overline{R}^2$	0.295	F-Statistic		50.694
DW	2.023	Prob(F-Stat.)		0.000
Observations	251	$\widehat{\sigma_u^2} \times 10^3$		0.018

Notes: Dependent variable is log differences of the rent index. DW is the Durbin Watson statistic.

Starting with

$$e_1^\top \left(\sum_{j=0}^{\infty} \rho^j \sum_{i=0}^{j} A^i \right) c$$

the constant in (3.33) is derived by observing that the double sum in the brackets is just

$$I(1 + \rho + \rho^2 + \ldots) + \rho A(1 + \rho + \rho^2 + \ldots) + (\rho A)^2(1 + \rho + \rho^2 + \ldots) + \ldots$$

and thus

$$\frac{1}{1-\rho} e_1^\top \left(\sum_{j=0}^{\infty} (\rho A)^j \right) c \, ,$$

which eventually gives the constant.

We obtain ρ by using rent receipts of 5080 apartment houses sold in Berlin during 1980 and 2000. We check the sensitivity of ρ with respect to plausible figures of administration and maintenance costs. According to these figures, the inverse capitalization rate θ lies between 0.39% and 0.44%. Rounding to the third digit, we obtain in all cases $\rho = 0.996$. Using the cap rate calculated with data of the RDM, and assuming maintenance cost of about 25% (30% for old buildings), we also obtain monthly cap rates in that range. The reference value in period t is then calculated as the sum of d_t^0 plus (3.33), see (3.6), where the unknown coefficients are replaced by their estimates.

We use the delta method for estimating the variances of the $\hat{p}_t^0$s, see Greene (2000, p. 298). Let $\delta \stackrel{\text{def}}{=} [\delta_0, \delta_1, \delta_2]^\top$ denote the vector of unknown coefficients from equation (3.31) and recall that p_t^0 is a function of these coefficients. Using a first order Taylor approximation of $p_t^0(\hat{\delta})$—$\hat{p}_t^0$ for short—around the true coefficient vector δ and calculating its variance gives

$$\mathcal{V}[\hat{p}_t^0] \approx \nabla p_t^0(\delta)^\top \mathcal{V}[\hat{\delta}] \nabla p_t^0(\delta) ,$$

where

$$\nabla p_t^0(\delta) = \begin{bmatrix} \dfrac{\partial p_t^0(\delta)}{\partial \delta_0} \\ \dfrac{\partial p_t^0(\delta)}{\partial \delta_1} \\ \dfrac{\partial p_t^0(\delta)}{\partial \delta_2} \end{bmatrix}$$

is the gradient of $p_t^0(\delta)$ and $\mathcal{V}[\hat{\delta}]$ is the covariance matrix of the OLS estimator $\hat{\delta}$. The variance of $\hat{p}_t^0$ is then estimated by computing the gradient at $\hat{\delta}$ and by replacing the covariance matrix with its estimate.

3.6.3 Market indicators

Here we present the data which are used to construct the market indicators in the return equation. Test for unit roots are conducted with the Augmented Dickey-Fuller (ADF) test. The tests are conducted with XploRe's `adf` Quantlet, which renders critical values proposed by MacKinnon (1991). The null of a unit root is rejected for all indicator series at commonly used significance levels. Table 3.5 reports the results plus summary statistics.

Unexpected inflation: Inflation rates are given by the log differences of the monthly consumer price index for four person households with average income in Berlin West. Data are provided by Berlin's Statistical Office (StaLa) in its Statistical Report M I 2. Unexpected inflation are the residuals of a AR(12) regression of the inflation rate on lagged values.

Ratio of building permissions: Log ratio of building permissions for single-family houses to building permissions for residential dwellings for Berlin West. Data are provided by Berlin's Statistical Office (StaLa) in its Statistical Report F II 1.

Table 3.5: *Summary statistics for the market indicators and ADF tests for 1982:8 to 1999:12.*

	Mean	Std. Dev.	$t(\hat{\rho})$	k_{max}	T
Unexpected inflation	0.000	0.002	-14.18^{*}	0	209
Ratio of building permissions	-0.447	0.315	-2.98^{**}	15	209
Spread10	0.009	0.001	-4.05^{*}	2	208
CDAX return	0.009	0.052	-12.24^{*}	0	209

Notes: $t(\hat{\rho})$ is the test statistic of the Augmented Dickey-Fuller (ADF) test for a unit root. * (**) indicates significance at the 1% (5%) level using MacKinnon's critical values for rejection of the hypothesis of a unit root. A constant is included and k_{max} is the maximal lag of included lagged differences. For the log ratios a parsimonious specification is chosen with $k = 1, 2, 3, 15$. T is the number of observations included in the regression.

Spread10: Difference between the monthly mortgage rate with interest rate fixation of ten years and returns on bonds of banks with a maturity of ten years. Data are provided by the Deutsche Bundesbank as SU0046 and WU8616. Mortgage rates are only available since 1982:6.

CDAX return: Stock Index CDAX reported by the Deutsche Börse AG which represents the average market trend of all publicly traded German cooperations. Data are provided by the Deutsche Bundesbank as WU001A. Monthly returns are calculated as the log differences.

Chapter 4

MD*Immo—Online Prediction of Berlin Single-Family House Prices

4.1 Introduction

> "It is somewhere to live; but a home is also, for many folk, a valuable asset. No wonder people love talking about house prices over the dinner table."
> The houses that saved the world, in: *The Economist*, March 30th 2002, p. 11

The previous Chapter 3 has presented a statistical model of single-family house prices in Berlin. The estimation results have shown that transaction prices are explainable at about 80% by the characteristics of the transacted house. This chapter describes the online price prediction service MD*Immo, which is based on the above given statistical model of single-family house prices.

MD*Immo, www.md-immo.com, is a cooperation between the CASE – Center for Applied Statistics and Economics at the Humboldt-Universität zu Berlin and the Office of Berlin's Surveyor Commission for Real Estate (GSt). It is MD*Immo's purpose to deliver reliable and prompt price predictions for single-family houses given the current state of the market and given the characteristics of the respective property.

Different to expert-based price indices by real estate bodies, see Section

2.3.3, MD*Immo gives nearly realtime price predictions and considers individual characteristics which lead to rebates or surcharges to the general price trend. It is the main goal of MD*Immo to increase the transparency in Berlin's real estate market for single-family houses.

The usual way for a potential house buyer or seller to collect information on the current state of the market consists in consulting ads for properties in newspapers. The most severe problem with this kind of information gathering is that ads only report offer prices. Real transaction prices will be lower, as the empirical results of Merlo and Ortalo-Magné (2002) suggest. Another possibility for a potential market participant to gain information on the market is to contact a real estate agent or a surveyor. However, this way to gather information will be expensive, especially, if the potential market participant needs the information only to decide if he will show up at the market at all. For this decision, MD*Immo is an interesting alternative for information gathering because

- its predictions are based on previously observed transaction prices
- online requests are up to now costfree
- it is easy to understand and to use.

This chapter shows how MD*Immo's price predictions are calculated, explains the necessary updating steps and presents its technical implementation. Section 4.2 presents the data used for fitting the statistical model and describes the updating process for the data. Moreover, the calculation procedure of the predictions are discussed. Section 4.3 explains the request process and the technical implementation. The last Section 4.4 gives an outlook on MD*Immo's future development.

4.2 Calculation of the price predictions

4.2.1 Time structure and data

The data for fitting the statistical model comprise all transactions of single-family homes in Berlin since the year 1996 up to the respective current date. The data are delivered by the GSt from its Automated Transaction Price Data Base (Automatisierte Kaufpreissammlung, AKS). The AKS comprises information on all transactions of single-family houses in Berlin like the price, the size of the lot, the age, and many variables which describe the condition of the structure and the location of the site. MD*Immo uses data since 1996 because the GSt introduced in that year some new variables which have to be reported for all transactions. These new variables include information on the location of the property which is an important characteristic. The data are used to fit the statistical model which is described in Chapter 3.

Every three months the GSt delivers data of newly incoming transactions and so every three month the statistical model is refitted for the respective extended data set. We introduce some notation to explain this process in detail. Let T^+ denote the month in which we obtain new data. Given that the GSt needs time to obtain all information on current sales, the newest transactions are for sales about three months ago. Let T denote the month with the newest transactions, the 'information lag' of our data set is about $T^+ - T = 3$ months. The market indicators have a lesser information lag, which is about one month. They are provided by the Deutsche Bundesbank and Berlin's Statistical Office.

MD*Immo uses the newest obtainable market indicators to calculate the price predictions and we do not use one-month forecasts for the market indicators. This makes the calculation procedure easier. Moreover, this simplification can be justified by the fact that the common price component—which is influenced by the market indicators—is rather sluggish, see the empirical results in Chapter 3. However, this means that MD*Immo transaction price predictions are lagged by one month. For example, if a user of MD*Immo requests a price prediction in March, then the prediction is calculated with the newest market indicators, which will be in this case from February. Thus, this price prediction is in effect a prediction of the transaction price for the property if it would have been sold in February.

4.2.2 Price predictions

The first step in calculating the price predictions consists in predicting the state vectors. The second step consists in predicting the log price for a single-family house given its characteristics, and the last step consists in retransforming the predicted log price into the price.

Recall from Section 3.3 that the state vectors contain the deviations of log prices from the reference value, the constant, the constant coefficients for the market indicators, and the implicit characteristic prices. Thus, in the first step the future expected return rate deviations are predicted, which apply to all house prices. The second step adds evaluated characteristics of the respective single-family house to the price prediction.

Let H denote the prediction horizon starting in month T and $h \in \{1, \ldots H\}$ the single prediction steps. Because the lag of the financial indicators is one month, we obtain that $2 \leqslant H \leqslant 4$. We have $H = 2$ for T^{+}, in which we obtain new data and $H = 4$ for last month before we obtain new data. Let the vector s_{T+h} collect the market indicators. Given the estimated $\hat{\phi}$ and $\widehat{\sigma_{\xi}^{2}}$, we can array the system matrices T_{T+h} and R of our SSM for Berlin single-family house prices, see equation (3.16a).

Starting with the last smoothed state vector $a_{T|T}$ and its covariance matrix $P_{T|T}$ from the fitted SSM, we have the following recursions for the predicted state vectors (Moryson; 1998, p. 25)

Start in $h = 1$:

$$a_{T+1|T} = T_{T+1} a_{T|T} \ , \quad P_{T+1|T} = T_{T+1} P_{T|T} T_{T+1}^{\top} + R$$

Step in $1 < h \leqslant H$:

$$a_{T+h|T} = T_{T+h} a_{T+h-1|T} \ , \quad P_{T+h|T} = T_{T+h} P_{T+h-1|T} T_{T+h}^{\top} + R$$

We need always the newest predictions of $\{a_{T+h|T}\}_{h=1}^{H}$ and $\{P_{T+h|T}\}_{h=1}^{H}$ for calculating the transaction price predictions.

For the second step in the log price prediction, we need the respective log reference value p_{T+H}^{0}, $\widehat{\sigma_{\varepsilon}^{2}}$, the transformation coefficients λ_j for the three continuous variables (lot size, floor space, and age), the transformation constants a_j and the standard deviations of the continuous variables. The latter three specifications are necessary for the transformation of the continuous variables. Moreover, let

$$z_n = \begin{bmatrix} -1 & 0_{K_\gamma} & 1 & T_\lambda(\text{lot size}) & T_\lambda(\text{floor space}) & T_\lambda(\text{age}) & d & 0 \end{bmatrix}$$

denote the vector that collects all information on the property for which the prediction is requested. The continuous variables are transformed according to the Box-Jenkins type transformations given in Section 3.3. The vector d contains indicators for discrete characteristics of the property. An indicator for a characteristic is 1 if the property has the characteristic and zero otherwise. The last vector in z_n contains indicator variables for characteristics which are included in the hedonic regression, but which are not at choice for the user. However, by definition, all of these non-selectable variables are coded in such a way that their respective indicator is zero for the most common specificity.

Recalling the definition $\Delta^0 p_{n,t} = p_{n,t} - p_t^0$ from the previous chapter, we have

$$p_{n,T+H} = p_{T+H}^0 + \Delta p_{n,T+H}^0 \ ,$$

so that the expected log price and the variance of the log price are given by

$$\begin{aligned} \mathcal{E}_T[p_{n,T+H}] &= p_{T+H}^0 + \mathcal{E}_T[\Delta p_{n,T+H}^0] \\ &= p_{T+H}^0 + z_n a_{T+H|T} \\ \mathcal{V}_T[p_{n,T+H}] &= z_n P_{T+H|T} z_n^\top + \sigma_\varepsilon^2 \ . \end{aligned} \tag{4.1}$$

Here we have used that

$$\begin{aligned} \mathcal{E}_T[\Delta p_{n,T+H}^0] &= z_n a_{T+H|T} \\ \mathcal{V}_T[\Delta p_{n,T+H}^0] &= z_n P_{T+H|T} z_n^\top + \sigma_\varepsilon^2 \end{aligned}$$

(Moryson; 1998, p. 25). Moreover, we have neglected that the reference value might be estimated. However, we found for the periods since 1996 that the growth rate of the rent index behaves like white noise. In that case, the reference value is given by the observed rent index and has not to be estimated.

Given z_n, p_{T+H}^0, and replacing σ_ε^2 by its estimate, the expressions for the expected log price and the respective variance from equations (4.1) can be calculated.

However, users of MD*Immo are not interested in log prices but in the prices itself. It follows immediately from Jensen's inequality that

$$\mathcal{E}[\exp(p_{n,t})] \neq \exp(\mathcal{E}[p_{n,t}])$$

so that a simple retransformation of the expected log price gives a biased estimate of the expected price. When the log prices $p_{n,t}$ are normally distributed, then the exponential transformations are log normally distributed and (Johnson and Kotz; 1970)

$$\mathcal{E}[\exp(p_{n,t})] = \exp(\mathcal{E}[p_{n,t}] + 0.5\sigma_\varepsilon^2) \tag{4.2}$$

$$\mathcal{V}[\exp(p_{n,t})] = \{\mathcal{E}[\exp(p_{n,t})]\}^2\{\exp(\sigma_\varepsilon^2) - 1\} \tag{4.3}$$

When the log prices are not normally distributed, the following approximation can be used to calculate the expectation and the variance of the exponential transformation. Using a second order Taylor approximation, we obtain

$$\begin{aligned}\exp(p_{n,t}) &\approx \exp(\mathcal{E}[p_{n,t}])\{1 + (p_{n,t} - \mathcal{E}[p_{n,t}])\} \\ &\quad +0.5\exp(\mathcal{E}[p_{n,t}])(p_{n,t} - \mathcal{E}[p_{n,t}])^2 \,.\end{aligned}$$

Using the second order approximation, one derives that

$$\mathcal{E}[\exp(p_{n,t})] \approx \exp(\mathcal{E}[p_{n,t}])\{1 + 0.5\sigma_\varepsilon^2\} \tag{4.4}$$

and using only the first order approximation, one obtains moreover that

$$\mathcal{V}[\exp(p_{n,t})] \approx \{\exp(\mathcal{E}[p_{n,t}])\}^2\sigma_\varepsilon^2 \,. \tag{4.5}$$

Both expressions can be calculated with the second expression given in (4.1) and an estimate of σ_ε^2. It is easy to prove that (4.4) can likewise be seen as first order approximation of (4.2). Given that fact, the expression for the expected price (4.4) is used for calculating the predicted price for a house with characteristics collected in z_n. The expression for the price variance (4.5) is used to calculate the standard deviation. These formulas are implemented in MD*Immo and are calculated online if a user requests a price prediction. The technical implementation of this process is described in the next section.

4.3 Technical implementation

The request process for a price prediction from www.md-immo.com consists of two steps. First, the user has to give information on the characteristics of the subject property. After sending the request, MD*Immo calculates in the second step the price prediction and the standard deviation. Then MD*Immo renders the results on the screen.

The request process is presented schematically in figure 4.1. It shows that currently a user of MD*Immo has to give the following information on the characteristics of the property for which he requests a price prediction:

- type of the property, which are row house, semi-detached house, and detached house
- district, where the user can choose between the new districts of Berlin according the reform in 2001

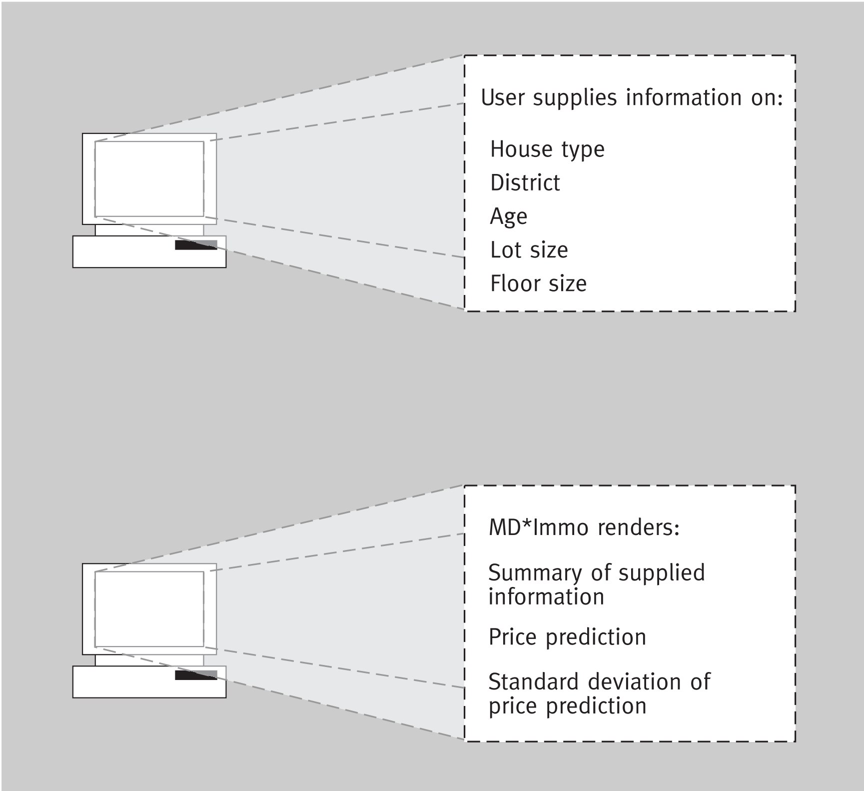

Figure 4.1: Input and output for a price prediction from MD*Immo.

- age of the property in years or, substitutional, year of construction
- size of the lot in square meters
- floor space in square meters.

Figure 4.2 shows a screenshot of MD*Immo's request page. Here, the type is a row house which is located in the district Spandau. The property is 52 years old, has a lot size of 255 square meters and a floor space of 158 square meters.

The technical implementation of MD*Immo consists of two components, which are the internet and the data processing component.

The internet component allows the interaction between the user and MD*Immo. It consists of static and dynamic HTML pages, where the dy-

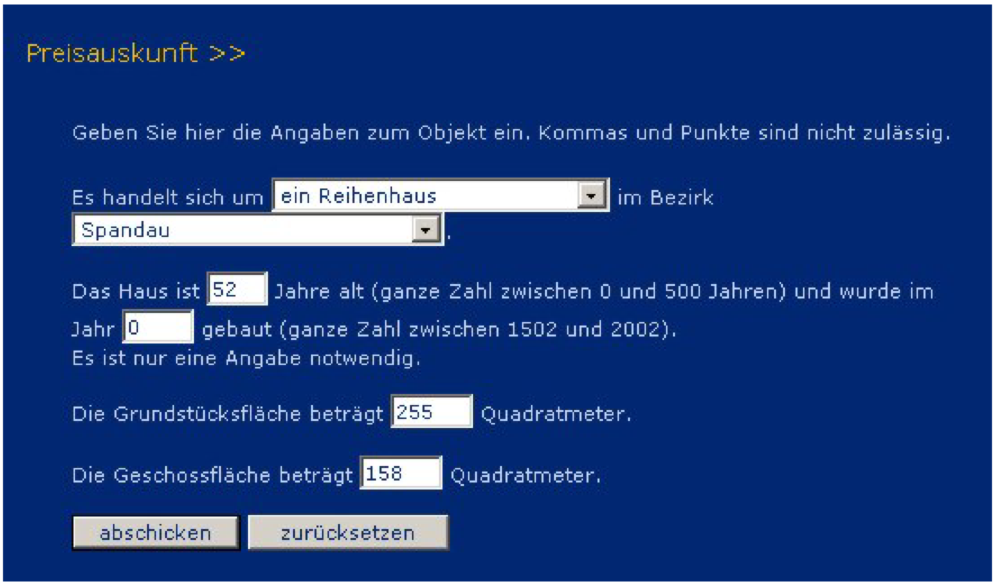

Figure 4.2: Screenshot of MD*Immo's request form, where the user has to provide the characteristics of the subject property.

namic pages are implemented as Java Server Pages. The validity of the requested characteristics are immediately checked for its consistency by JavaScript. Negative numbers for the age and the sizes are ruled out as well as letters or special characters. Dialog boxes inform the user about incorrectly specified characteristics. If the requested characteristics are valid, the request can be transferred.

After receiving the information on the characteristics, several matrix operations are used—together with the current estimated coefficients—for calculating the price prediction according to the formulas (4.1), (4.4), and (4.5).

MD*Immo uses the Java Server Page web server Tomcat as well as the Java matrix library JAMA for running these calculations. Moreover, as a part of the internet component, the requested characteristics of the subject property are saved in MD*Base, www.mdtech.de. These data will be used for subsequent statistical analysis of the requests.

The data component consists of data delivery by the GSt, the estimation of the statistical model and the calculation of the state predictions, see the previous section. After this is done, the estimated coefficients are imported via SQL statements into MD*Base.

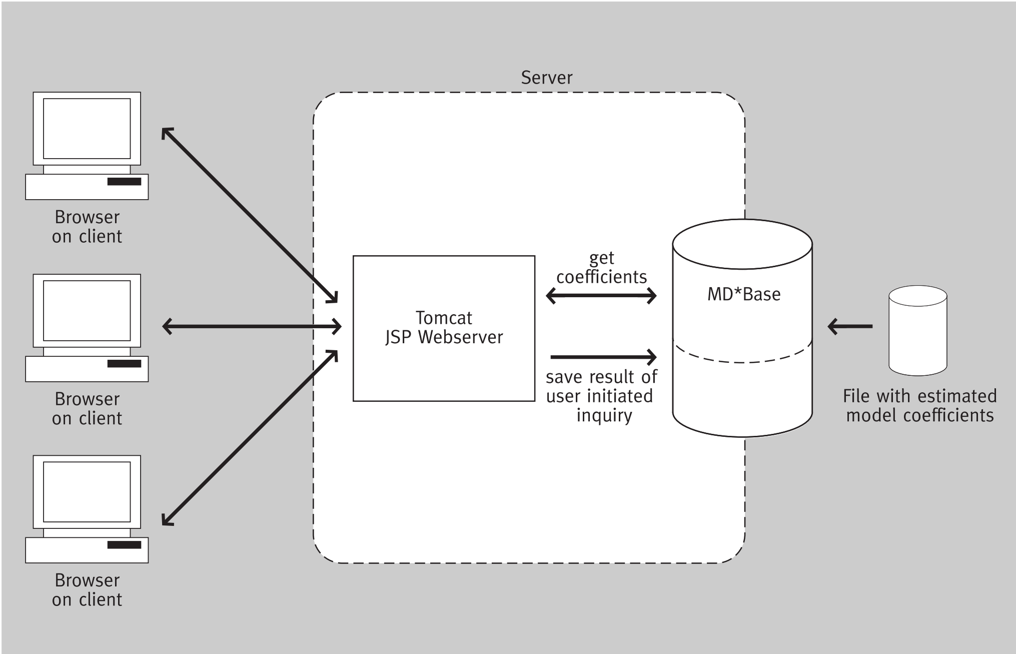

Figure 4.3: MD*Immo's technical implementation. It consists of an internet and a data component.

The two components of the technical implementation are show schematically in figure 4.3. The data component is shown on the right and the internet component on the left hand side. MD*Base serves as interface for the internet component and connects both components.

4.4 Outlook

The future of MD*Immo looks promising. On behalf of Berlin's Surveyor Commission (GAA), the CASE – Center for Applied Statistics and Economics at the Humboldt-Universität zu Berlin will implement MD*Immo as new a product of GAA Online, the web service of the GAA.

Currently, users of GAA Online can request average prices from transacted properties in the vicinity of the subject property, i.e., the property for which a price prediction is requested. Average prices do not explicitly consider the heterogeneity of properties and may be misleading if the subject property is only weakly comparable to the transacted properties, see the discussion in Section 2.4. Moreover, if no current transactions are available in the database (AKS) of the Surveyor Commission, GAA Online calculates no average price. In this case, the service renders a message which informs the user that no comparable sales are available. Furthermore, GAA Online uses historical data to calculated average prices, meaning the present state of the market—the common price component—is not considered.

MD*Immo's online service will enhance the offerings of GAA Online in several respects. Characteristics are explicitly considered in MD*Immo's hedonic model and the estimated implicit prices can value—in principle—every requested combination of characteristics. Nevertheless, queries are not possible for characteristics that have not been observed for historical transactions. For example, without sales from a specific neighborhood, it is impossible to estimate an implicit price for that neighborhood. It is possible that no sales are observed because there exist no single-family houses, for example, because it is an inner-city neighborhood. Therefore, missing characteristics due to non observed sales do not seem to be a severe problem. Different to GAA Online, MD*Immo considers directly the current state of the market, because it explicitly models the behavior of the common price component.

The GAA will deliver extended data for fitting MD*Immo's statistical model. The greatest novelty is the identification of each location in Berlin by univocal location coordinates. In its current state, the value of a location is treated rather unsophisticatedly via district dummies, which might be too crude to capture the whole location value of a property. This value will depend on amenities in the neighborhood, the distance to the city and employment centers, see the model of location value in Section 2.2. The coordinates may capture the location value better than simple district dummies. Furthermore, given that the location coordinates will be significant in an extended hedonic model, the request form can be improved by explicitly considering the street address of a property. This will increase the usability of MD*Immo's online prediction service.

It is the main goal of CASE and GSt to increase the transparency of Berlin's market for single-family houses. Via online services, potential buyers and sellers can request price predictions for their subject property and will be informed about the current state of the market. This information will serve as an orientation for their investment decisions. A successful implementation of MD*Immo via GAA Online will increase the acceptance of valuation techniques that are based on large data sets and statistical techniques. Although such techniques are interesting for estimating adjustment factors for the sales comparison approach—see the discussion in Section 2.1.2—up to now German valuers make no use of such techniques.

Chapter 5

An Empirical Analysis of Income Valuation in Germany

5.1 Introduction

> "A rule of thumb, artfully employed, sometimes beats a complex discounted cash flow calculation hands down." Brealey and Myers (2000, p. 82)

Real estate valuation is the task of appraising the prospective price of a site or building in the case of a sale. Such appraisals are important for investment decisions, for real estate funds and project developments. Additionally, real estate performance indexes like the German property index DIX are calculated with appraisals, see Section 2.3.3. For appraising income generating properties, the income approach is commonly used, which discounts expected future income to obtain an appraisal for the subject property.

Many European countries have own valuation techniques based on the income approach (McParland et al.; 2002). The internationalization of real estate suggests an investigation of such techniques, because foreign investors need to understand the concepts that national appraisers use. In addition to the need of clarifying existing valuation concepts, there is also a need for a better empirical assessment of the accuracy of property valuations, see the statements of the *Carsberg Report* from the RICS (2002). Only such assessments allow to evaluate the risk inherent in competing national techniques.

Crosby (2000) gives a detailed overview on studies that assess the accuracy of appraisals from Anglo-American countries. In addition to studies that compare transaction prices and appraisals, there exist also studies that compare different appraisals for identical objects, see Graff and Young (1999). One can criticize such studies because they implicitly assess appraisers and not valuation techniques. That critique is not applicable to our study. Unlike other European countries, appraisers in Germany are obliged in many cases to use codified valuation techniques, see Rüchardt (2001) and Section 2.1.6.

In this chapter, we clarify German valuation techniques of the income approach and assess its outcomes with 4150 observations for apartment houses in Berlin. We explain short-run deviations between prices and appraisals by incompletely appraised object-specific characteristics and by market indicators. Due to the fact that valuation according to the income approach is not restricted to properties, this study is of general interest for appraisers, banks and investors, if they need valuations of real estate, stocks or companies.

The chapter proceeds as follows. Section 5.2 clarifies the rationale of the income approach and presents valuation techniques. In Section 5.3, economic theory is used to derive implications for the relationship between transaction prices and appraisals. Section 5.4 presents the data and explores the accuracy of the valuation techniques. Section 5.5 explains the short-term deviation between prices and appraisals by market indicators. The final section concludes.

5.2 Income approach

The economic rationale of the income approach for existing properties is that no investor will pay more for a property than he will retrieve by holding the property. Every month, the owner of an income generating property receives the rents of the tenants. A part of the rents provide cover for operating costs and the income that remains after subtracting all costs is called net operating rent.

Let t denote the current period and consider that property n can be utilized up to period $T_{n,t}$, so that the property has a wreckage value in

$$\tau_{n,t} \stackrel{\text{def}}{=} T_{n,t} - t$$

periods. With no demolition costs, the lower bound of the wreckage value is just the value $L_{n,t}$ of the land. Given rational investors, the value $V_{n,t}$ of the property in period t is its present value

$$V_{n,t} = \mathcal{E}_t\left[\sum_{j=1}^{\tau_{n,t}} \frac{D_{n,t+j}}{\prod_{i=1}^{j}(1+R_{t+i})}\right] + \mathcal{E}_t\left[\frac{L_{n,t+\tau_{n,t}}}{\prod_{i=1}^{\tau_{n,t}}(1+R_{t+i})}\right] . \tag{5.1}$$

Here, $\mathcal{E}_t[\cdot]$ denotes the expectation operator given all information up to period t. D_{t+j} is the net operating rent for the property in period $t+j$ and R_{t+j} are discount rates for income in that period. The discount rates are returns that investors require for investments in real estate and are equal to returns of other investments that share the same risk.

Reasoning that the land value will be given by the discounted net operating rents for all subsequent buildings on the site, we obtain

$$V_{n,t} = \mathcal{E}_t\left[\sum_{j=1}^{\infty} \frac{D_{n,t+j}}{\prod_{i=1}^{j}(1+R_{t+i})}\right] . \tag{5.2}$$

Thus, the current value of the property is given by the whole discounted expected income it will generate in the future.

In the practice of valuations according to the income approach, simplified versions of the present value (5.1)—respectively (5.2)—are used, see Brown and Matysiak (2000) and Section 2.1.3. Firstly, often the time-varying discount rates are replaced by its long run average R and secondly, a constant growth rate of the net operating rents is assumed. Both simplifications underly valuation techniques according to the income approach in Germany.

5.2.1 Valuation according to WertV

In Germany, real estate valuation is codified through the Regulation on Valuation (WertV) and the Guidelines on Valuation (WertR 91). The central figure in the WertV is the market value (Verkehrswert). It is the transaction price one should expect for a property given its characteristics, the general market conditions and given usual business dealings (§ 194 BauGB). For income generating properties, the market value is thus equal to the present value $V_{n,t}$.

The determination of the market value of an income generating property according to WertV follows a two-step procedure. In the first step, the income value (Ertragswert) of the respective house has to be determined. In the second step, the calculated income value has to be adjusted for general market conditions to derive the ascertained market value (§ 7 WertV). The market conditions are influenced by the current situation of the economy, by financial conditions and special conditions of the respective region (§ 3 Abs. 3 WertV).

The income value $E_{n,t}$ is (§§ 15, 16 WertV)

$$E_{n,t} = \frac{1}{\theta_t} \left\{ 1 - \left(\frac{1}{1+\theta_t} \right)^{\tau_{n,t}} \right\} (D_{n,t} - \theta_t L_{n,t}) + L_{n,t} \ , \tag{5.3}$$

where four figures are needed: the expected lasting net operating rents $D_{n,t}$, the remaining time of usage $\tau_{n,t}$, the value of the site $L_{n,t}$ and the discount rate θ_t. The appraiser assesses $D_{n,t}$ and $\tau_{n,t}$ according to the rules in WertR 91, given information on the property, and given his knowledge of the market. The value of the site $L_{n,t}$ (Bodenwert) has to be determined by using transaction prices of comparable sites (§ 13 Abs. 1 WertV). If that is not possible, the appraiser must use approximate values. Such values are delivered by the regional Surveyor Commissions for Real Estate, which also provide figures for θ_t (Liegenschaftszinsen). These yields are averages of internal rates of return calculated with (5.3) after replacing $E_{n,t}$ by observed historical transaction prices, see Gottschalk (1999, B III) and § 11 Abs. 2 WertV. The presently effective rates for apartment houses in Berlin are calculated with historical data from 1996 up to 1999 and replace the discount rates that were calculated in 1996 (Senatsverwaltung für Stadtentwicklung; 2000).

The above given formula is a simplification of the present value. Reformulation of (5.3) gives

$$E_{n,t} = \frac{1}{\theta_t} \left\{ 1 - \left(\frac{1}{1+\theta_t} \right)^{\tau_{n,t}} \right\} D_{n,t} + \left(\frac{1}{1+\theta_t} \right)^{\tau_{n,t}} L_{n,t} \ . \tag{5.4}$$

One obtains exactly this expression from the present value (5.1) by assuming a constant discount rate R, a constant expected growth rate $G < R$ of net operating rents and land values for all future periods and by defining

$$\theta \stackrel{\text{def}}{=} \frac{R-G}{1+G} \ .$$

Two remarks are in order: first, the above derivation of the income value from the present value suggests that the discount rate should be estimated as the long-run average internal rate of return. Thus, it is questionable why the GAA uses only observations from four years to estimate this rate. Second, the usage of $L_{n,t}$ in (5.4) might be justified by the argument that it gives the current value of the site with the option of optimal development (Capozza and Li; 2002). If the current building is not optimal given its location and usage, $(1+G)^{\tau_{n,t}} L_{n,t}$ is a plausible guess for the value of the site when the existent building is used up. In standard real options models, the value of the development option—here the value of the land—will grow at least with the rate of the net operating rents (Pindyck; 1991; Dixit and Pindyck; 1994). Thus, $(1+G)^{\tau_{n,t}} L_{n,t}$ is a conservative guess.

Whereas the determination of the income value is codified in detail, the market adjustment is open for the judgement of the appraiser. In many cases, valuers use proportional correction factors that are calculated by professional bodies or Surveyor Commissions, see Gottschalk (1999, C V). Denoting M_t the adjustment factor for general market conditions, the ascertained market value is

$$V^a_{n,t} = M_t E_{n,t} \ . \tag{5.5}$$

In effect, the two-step procedure divides the market value into two components. The first component gives the value of the property calculated by using a constant discount factor. The second step adjusts for the fact that the returns that investors require might be currently higher or lower than they will be in the long run.

5.2.2 Valuation according to the multiple technique

In addition to income valuation according to WertV, German appraisers use the simpler multiple technique, which is also known as the real estate agent method (Maklermethode). According to this technique, the ascertained market value of property n is just the multiplied current gross rent

$$V^a_{n,t} = c^g D^g_{n,t} \ . \tag{5.6}$$

Gross rents contain distributable operating costs like land-tax, cleaning services, insurances, charges of public utilities, which are passed through to the tenants and non-distributable operating costs like management and maintenance costs. We obtain exactly the above expression from the present value (5.4) by assuming a constant discount rate R, a constant expected growth rate G of net operating rents, constant relative operating costs O, and by defining

$$\theta^g \stackrel{\text{def}}{=} \frac{(1+O)(R-G)}{1+G}$$

and

$$c^g \stackrel{\text{def}}{=} \frac{1}{\theta^g} \, .$$

The capitalization factor c^g is often provided by real estate professional bodies, calculated with information given by its members or by using historical prices and gross rents. Thus, the multiple technique is a rule of thumb, that needs only two figures for providing an appraisal of the market value. No additional information has to be gathered than the gross rent figure and the corresponding capitalization factor.

However, as Gottschalk (1999, C VII) emphasizes, the assumption of constant relative operating costs is critical because it ignores property-specific operating costs, discounts due to access rights or rights to way, and conditions of the property that may influence lasting rents. All of these facts have to be considered explicitly in income valuation according to WertV.

5.3 How to evaluate different valuation techniques?

Difficulties of evaluating different valuation techniques arise because economic loss functions are context-sensitive and do not always coincide with statistical forecast evaluation measures (Diebold and Mariano; 1995). This point is easily seen for valuation techniques that are used for appraising the collateral value of properties for lending purposes. A technique that undervalues on average might be preferable to a technique that is unbiased, but exhibits outliers that overestimate the true collateral value at a high degree. Whereas the first technique is too conservative, because it rejects some good applicants, the second technique generates high losses in the case of default. Stated an asymmetric loss given default function, the first technique will be economically preferable (Shiller and Weiss; 1999).

Income valuation, however, is used for many different purposes and the associated loss functions will exhibit non-comparable features. Thus, we have to concentrate on accuracy intentions inherent in WertV. In addition, we provide various criteria which are valuable for investors given their purposes. Clapp and Giaccotto (2002) argue that estimated distributions are good graphical devices for comparing different valuation techniques. Investors can apply their own weights to appraisal errors for choosing their preferred valuation technique.

5.3.1 Statistical model based on WertV

Let $P_{n,t}$ denote the transaction price of property n in period t, then

$$P_{n,t} = V_{n,t}U_{n,t} \ , \tag{5.7}$$

where $U_{n,t}$ is the unsystematical component and $\mathcal{E}_t[U_{n,t}] = 1$. Unsystematic deviations may happen if one of the contracting parties has a special interest in obtaining the object or if there are personal relationships between seller and buyer. Multiplicative disturbances are justified by the fact that proportional figures are common in real estate business. For example, fees of real estate agents are proportional to prices and valuers—according to personal communication—often use proportional discounts or surcharges, see also § 14, 25 WertV. The expected transaction price is equal to the market value

$$\mathcal{E}_t[P_{n,t}] = V_{n,t} \ .$$

Replacing the market value in (5.7) with the ascertained market value $M_tE_{n,t}$ and defining the ratio between price and income value as

$$Q_{n,t} \stackrel{\text{def}}{=} \frac{P_{n,t}}{E_{n,t}} \ , \tag{5.8}$$

one obtains

$$Q_{n,t} = M_tU_{n,t} \ . \tag{5.9}$$

According to WertV, it is guaranteed that $E_{n,t} > 0$, see appendix 5.7.

The general market condition M_t does not depend on n and is independent of the idiosyncratic unsystematical component $U_{n,t}$. The definition of the general market condition as adjustment factor for the current situation on the market implies that $\mathcal{E}[M_t] = 1$ and thus

$$\mathcal{E}[Q_{n,t}] = 1 \ . \tag{5.10}$$

So, the unconditional expected deviation between price and income value must be zero and both must coincide in the long run. However, for the short-run we have

$$\mathcal{E}_t[Q_{n,t}] = M_t \ . \tag{5.11}$$

The expected deviation between price and income value in the short run is given by the general market condition M_t.

Extracting this model from the WertV leads to two testable hypotheses: the average ratio of price to income value should be equal to one for a sample that covers several years. The income values calculated according to WertV must be unbiased on average, see (5.10). Although Berlin's Surveyor Commission reports Q in its data base, one might object that the reciprocal of Q is also of interest. Dotzour (1988) uses this figure for evaluating residential appraisal errors. From Jensen's inequality it follows that $\mathcal{E}[Q] > 1$ if $\mathcal{E}[1/Q] = 1$. Thus, we check also the hypothesis

$$\mathcal{E}[1/Q_{n,t}] = 1 \; . \tag{5.12}$$

Such ambiguities arise because we have no economic loss function at hand and it is not obvious which ratio is preferable. The second hypothesis is robust against such objections. It states that any deviation between prices and income values for single periods must be systematic and not explainable by characteristics of the respective houses, see (5.11). Additionally, we extend the above stated hypotheses by a third one: the deviations of the income values around the prices should be smaller than the deviations that are generated by the multiple technique.

5.3.2 What are the general market conditions?

The market value is the figure of interest in property valuation. However, whereas the technique for appraising the income value according to WertV is prescribed in detail, the adjustment for general market conditions is not. Is it a good idea to leave this important adjustment at the will of an appraiser?

Pretending that we could extract the general market conditions M_t, we can try to find reasonable economic indicators that explain the behavior of this component. By doing this, we can show that the idea behind the adjustment procedure outlined in the WertV is reasonable. Thus, we can not test if an individual appraiser uses the right indicators to assess the general market conditions. But we can test if the idea is a good one and if the indicators that are quoted in the WertV (§ 3) and in the Reports of the GAA (Geschäftsstelle des Gutachterausschusses für Grundstückswerte in Berlin; 2001) are reasonable.

For an explanation of the general market conditions, we use the log-linearized version of the present value (Campbell et al.; 1997; Cochrane;

2001).

$$\ln V_{n,t} \approx \sum_{j=0}^{\infty} \rho^j \left\{k + (1-\rho)\mathcal{E}_t[d_{n,t+1+j}] - \mathcal{E}_t[r_{t+1+j}]\right\} \ .$$

is a first order approximation of the market value (5.2). Here, $d_{n,t}$ denotes the logarithm of the net operating rents and $r_t \stackrel{\text{def}}{=} \ln(1+R_t)$. k and ρ are approximation constants with $0 < \rho < 1$. Given a constant return rate r, we suppose as approximation of the income value according to WertV

$$\ln E_{n,t} \approx \sum_{j=0}^{\infty} \rho^j \{k + (1-\rho)\mathcal{E}_t[d_{n,t+1+j}] - r\} \ . \tag{5.13}$$

Let $q_{n,t}$ denote the log ratio of price to income value (5.8). We obtain with (5.7) and the above stated approximations

$$q_{n,t} = \kappa - \sum_{j=0}^{\infty} \rho^j \mathcal{E}_t[r_{t+1+j} - r] + \varepsilon_{n,t} \ , \tag{5.14}$$

where the new constant κ guarantees that $\varepsilon_{n,t} \sim (0, \sigma_\varepsilon^2)$. We obtain for the log general market condition with (5.9)

$$\mathcal{E}_t[q_{n,t}] = m_t \tag{5.15}$$

and thus

$$m_t = \kappa - \sum_{j=0}^{\infty} \rho^j \mathcal{E}_t[r_{t+1+j} - r] \ . \tag{5.16}$$

Short-run deviations between prices and income values are deviations of the expected returns from its long-run average. These deviations might be influenced by proxies of risks that influence returns (Ling and Naranjo; 1997). We denote the risk proxies below as market indicators. Expected return deviations are modelled as

$$\mathcal{E}_t[r_{t+1} - r] = x_t \ ,$$

where x_t is a stationary process with

$$\alpha(L)x_t = s_t\gamma + \xi_t \ . \tag{5.17}$$

Here, $\alpha(L)$ is a polynomial in the lag operator L with $L^j x_t = x_{t-j}$. The row vector s_t comprises news in market indicators. ξ_t comprises unsystematical influences on the expected return deviation. We obtain for the sum of discounted expectations

$$-\sum_{j=0}^{\infty} \rho^j \mathcal{E}_t[r_{t+1+j} - r] = \psi(L)\alpha(L)x_t \; , \tag{5.18}$$

where

$$\psi(L) \stackrel{\text{def}}{=} \frac{L\alpha(L)^{-1} - \rho\alpha(\rho)^{-1}}{\rho - L}$$

is a stationary lag polynomial (Gourieroux and Monfort; 1997, p. 473). Plugging this expression into (5.16) yields

$$m_t = \kappa + \psi(L)\alpha(L)x_t \; .$$

We obtain with (5.17)

$$\phi(L)m_t = \tilde{\kappa} + s_t\gamma + \xi_t \tag{5.19a}$$

with $\phi(L) \stackrel{\text{def}}{=} \psi(L)^{-1}$, $\tilde{\kappa} \stackrel{\text{def}}{=} \kappa\phi(1)$, and $\mathcal{E}[m_t] = \kappa$. Using (5.14) and (5.16), we obtain eventually

$$q_{n,t} = m_t + \varepsilon_{n,t} \; . \tag{5.19b}$$

The system of equations (5.19) resembles the two-step valuation procedure according to WertV: deviations between log prices and income values are due to currently prevailing general market conditions. We interpret the market conditions as short-run deviations in returns that investors require for investment in real estate which are influenced by market indicators.

5.4 Empirical investigation

5.4.1 Data

The main data set contains 4150 transaction observations of apartment houses from January 1980 to May 2000. The data are stored in the non-public part of MD*Base, www.mdtech.de, and is provided by Berlin's Surveyor Commission for Real Estate (GAA). Surveyor Commissions have to collect all relevant information on real estate transactions in its federal state (§§ 192-199

Table 5.1: *Number of observations with appraised income value per year.*

Year										
1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
251	197	193	382	335	253	265	287	393	315	236
Year										
1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	Total
158	115	110	107	70	62	78	212	105	26	4150

Note: Year 2000 comprises only observations up to May.

BauGB). Moreover, different indices of the Statistical Office Berlin (StaLa) and of the Deutsche Bundesbank are used.

The income values according to WertV are appraised by surveyors of the GAA for internal usage. If information on a transaction is insufficient, a questionnaire is sent to the owners of the property to gather the relevant figures. If the questionnaire is not returned, the respective appraisers impute figures from experience for the missing numbers. Most of these figures are explicitly specified in the WertR 91.

It might be problematic that appraisals are done after transactions have happened because it gives the valuer the opportunity to use ex-post knowledge to derive 'good' appraisals (Crosby; 2000). The GAA assured that they appraise 'mechanically' according to WertV and that information on transaction prices is not used. Two further facts strengthen this statement: deviations between income values and transaction prices are plausible due to the fact that income valuation is only the first step in ascertaining the market value. Furthermore, the valuations are for internal usage and no pressure is put on the objectivity by instructing clients.

Table 5.1 gives the number of observations with appraised income values per year. Before 1994, we have only transactions from the West part of Berlin. According to personal communication with the surveyors of the GAA, the transaction volume is mostly influenced by changes in taxes and subsidies. The number of appraised properties depends also on the completeness of the information needed for valuation. Other variables in our data set are gross or—respectively—net rents, age of the building, size of the floor space and size of the lot. Table 5.2 gives an overview on the variables. Here, the age of the building is the age at the time of transaction. Before 1995, it was common to report the yearly gross rent, since 1995 it is more common to report the

Table 5.2: *Summary statistics for transacted apartment houses in Berlin, Germany between 1980:1 to 2000:5.*

	Mean	Median	Std. Dev.	Min	Max	Units
Lot size	982.2	767.0	1920.8	186.0	56332.0	*Square metres*
Floor space	2168.9	1867.5	2637.2	128.0	89614.0	*Square metres*
Age	73.9	81	29.2	0	186	*Years*
Price	721.2	496.0	1120.9	53.7	40900.0	*Thsd. EUR*
Income value	662.2	455.6	1451.8	48.1	72800.0	*Thsd. EUR*
Gross rent	54.8	43.3	90.0	6.2	4260.5	*Thsd. EUR*
Net rent	61.9	33.6	131.0	3.6	1610.7	*Thsd. EUR*

Notes: Original currency units are German marks which are converted to EUR by dividing with 1.95583. 3835 observations have information on the gross rent and 315 on the net rent. Income values are calculated by the surveyors of the GAA according to WertV.

yearly net rent, that is the gross rent without distributable operating costs.

Figure 5.1 shows the average yearly price-rent ratios for different vintages. It reveals that the implied multipliers increased after the German Reunification in the early 90ies and felt back to its average of about 12.5 after 1995.

5.4.2 Assessment of valuations according to WertV

Table 5.3 reports summary statistics for ratios of price to income value and the inverse ratios for a time span of more than 20 years.

Panel A of Table 5.3 reports summary statistics for ratios of price to income value. It shows that the appraisal error is 13.3% and income values understate prices on average. The median appraisal error is smaller than the average error and amounts 5.8%. The positive skewness underlines that the density of ratios is not symmetric. The excess kurtosis of 16.178 reveals a leptokurtic density with more mass in the middle compared with a normal distribution (Spanos; 1999).

Panel B of Table 5.3 reports summary statistics for ratios of income value to price. The mean of the ratios is less than one. On average, the relative appraisal error is -3.2% and prices were understated. The median error of -5.5% is even larger. The positive skewness underlines that the density of ratios is not symmetric and the excess kurtosis of 9.147 reveals a leptokurtic density with more mass in the middle compared with a normal distribution.

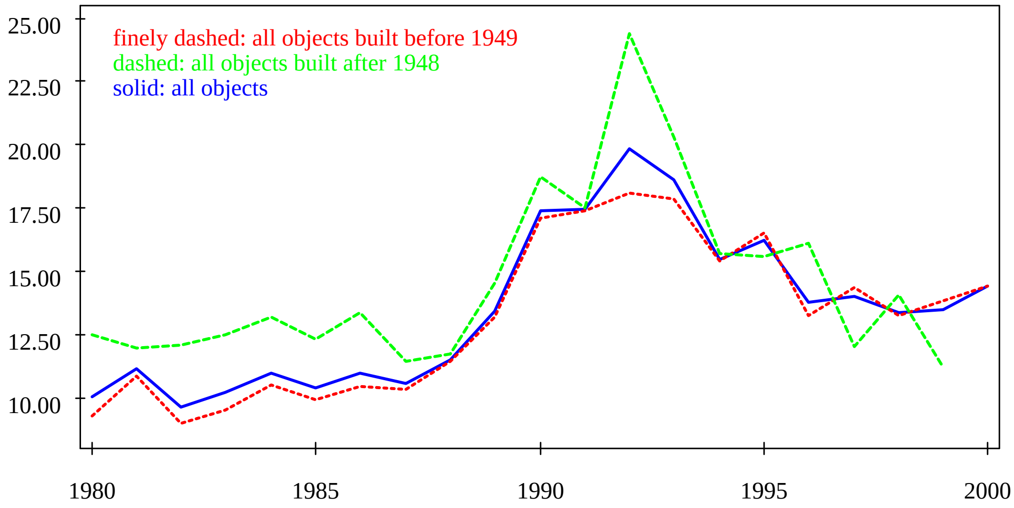

Figure 5.1: Average yearly price-rent ratios from 1980 to 2000. Calculated with all 3835 observations with information on gross rents.

For the hypothesis (5.10) that $\mathcal{E}[Q] = 1$, we obtain a t-statistic of more than 21.8 and reject that income values are unbiased predictors for transaction prices. For the hypothesis (5.12) that $\mathcal{E}[1/Q] = 1$, we obtain a t-statistic of -7.06 and reject once again that income values are unbiased predictors of transaction prices. Thus, income values according to WertV are smaller than prices on average, which holds irrespective of the ratio that is used for assessing unbiasedness.

The average deviations of about -3.2% for ratios of income values to prices appear to be large compared with the results of Dotzour (1988), who found an average appraisal error of 0.06% calculated with ratios of appraisal to price. Chinloy et al. (1997) found for US data, that appraisals are on average about 2% higher than prices. Their explanation for this result is that appraisers have an incentive to overappraise, because they are paid only after a successful deal.

As we have already stated, surveyors of the GAA are independent of any instructing clients and will have no incentive to over- or understate income values. However, if the information on a property is insufficient for appraising it, approximate figures have to be used which are reported in WertR 91.

Table 5.3: *Summary statistics for ratios of price to income value and for ratios of income value to price. Income values are appraised according to WertV.*

Panel A: Ratios of Price to Income Value				
Mean	Standard deviation	Minimum	Median	Maximum
1.133	0.392	0.247	1.058	4.941
10% Quantile	90% Quantile	Skewness	Kurtosis	Number of obs.
0.773	1.577	2.726	19.178	4150
Panel B: Ratios of Income Value to Price				
Mean	Standard deviation	Minimum	Median	Maximum
0.968	0.294	0.202	0.945	4.044
10% Quantile	90% Quantile	Skewness	Kurtosis	Observations
0.634	1.294	1.562	12.147	4150

Thus, we check if this may explain the appraisal errors.

Inspection of the income value (5.4) reveals that there are at least four possible explanations—individual and in combination—for an understatement (see appendix 5.7)

- the discount rate θ_t is too large on average and prospected net operating rents are discounted too much
- prospected net operating rents $D_{n,t}$ are valued too low on average
- values $L_{n,t}$ for lots are too low on average
- remaining times of usage $\tau_{n,t}$ are too low on average.

Incorrect rating of discount rates is a simple explanation for average understatement of prices when the rates are too high. Whereas the discount rate influences all appraised income values in the same way, the other three factors are specific to the house under valuation. Object-specific factors should have no explanatory power for deviations between price and income value. The log-linearized version of the hypothesis is given by

$$q_{n,t} = m_t + \varepsilon_{n,t} \ , \tag{5.20}$$

which is equivalent to (5.19b). We test for mis-adjustment by running a regression of the log ratios on unadjusted characteristics. Recall that

Table 5.4: *Linear regression for q on object-specific characteristics.*

	Coefficient		t-Statistic	P-Value
Log real gross rent	0.034		2.89	0.004
Log lot size	-0.027		-2.01	0.045
Age	-0.001		-4.89	0.000
		Diagnostics		
R^2	0.218	$\bar{R}^2$		0.166
F-Statistic	4.197	P-Value(F-Stat.)		0.000
Observations	3835	$\widehat{\sigma^2_\varepsilon}$		0.069

Notes: Coefficients for overall constant and time dummies are not reported. Gross rents are deflated with StaLa consumer price index for households comprising four persons with average income in Berlin West, base year is 1995.

$\ln(1/Q) = -q$ and so the qualitative results of the regression will not depend on whether $\ln Q$ or $\ln(1/Q)$ is the dependent variable. If net operating rents are set correctly, then gross rents should have no influence on q. Due to the fact that q is dimensionless, we deflate the gross rents with the yearly StaLa consumer price index. Given a correctly specified lot value, the size of the lot should have no influence on q. Eventually, when surveyors judge the remaining time of usage correctly, the age should have no influence on q.

Table 5.4 reports that all three unadjusted characteristics influence log deviation between price and income value. Monthly time dummies are included to cope for general market conditions. The qualitative results remain essentially unchanged if no dummies are included. However, the explanatory power decreases substantially, where $R^2 = 0.01$.

The positive elasticity for gross rents is explainable with the prescribed practice to adjust gross rents with constant relative operating costs when no property-specific information is available (Gottschalk; 1999, p. 278). This ignores that a part of the operating costs is not proportional to gross rents, like costs for housekeeping, cleaning, maintenance, and management. The negative influence of the lot size can be explained by the fact that lot values are appraised with the sales comparison approach. Such appraisals will be more reliable for normal-sized sites, where more sales are observed and more sales can be used for appraising the lot value. The reliability will decrease for larger sites, where less comparable sales are at hand. Eventually, it is plausible that the remaining time of usage is a decreasing function in the age of the building. Assuming that the remaining time of usage is assessed correctly for new buildings, then the negative coefficient suggests that the remaining time for older buildings is assessed too high.

5.4.3 Comparison with the multiple technique

We have shown that income values according to WertV are biased appraisals of prices. Nevertheless, concluding that valuations according to WertV are inaccurate is a little bit too hasty, at least if we have no better alternative. We will compare the outcomes of income valuation according to WertV with the outcomes of the simpler multiple technique.

According to the multiple technique, an appraisal for a property is derived by multiplying its gross rent with a capitalization factor, where the factor is one divided by the gross discount rate. Thus, the income value according to the multiple technique is $V^a = D^g/\theta^g$.

However, to use this appraisal formula, we need θ^g. We set the discount rates for our data so that average price to appraisal ratios are equal to one. Let θ_{PM} denote the discount factor that guarantees for $\overline{P/V^a} = 1$ and let θ_{MP} denote the discount factor that guarantees for $\overline{V^a/P} = 1$.

$$\theta_{PM} = \frac{\sum_{t=1}^{T} N_t}{\sum_{t=1}^{T} \sum_{n=1}^{N_t} \left(\frac{D^g_{n,t}}{P_{n,t}} \right)^{-1}}$$

is given by the harmonic mean of the ratios of gross rent to price. Here, t is the index for the periods and N_t is the number of observation per period. We obtain analogously

$$\theta_{MP} = \frac{1}{\sum_{t=1}^{T} N_t} \sum_{t=1}^{T} \sum_{n=1}^{N_t} \frac{D^g_{n,t}}{P_{n,t}}$$

as the arithmetic mean of the ratios. Calculating these figures with all 3835 observations with information on gross rents yields $\theta_{PM} = 8\%$ and $\theta_{MP} = 9.17\%$. The corresponding capitalization factors for gross rents are 12.5 and 10.9. An accurate valuation technique should be unbiased, but it should also have a small variance. A convenient measure for evaluating this relation is the mean squared prediction error (MSPE)

$$\mathcal{E}\left[(X-a)^2\right] = \mathcal{V}[X] + \left(\mathcal{E}[X]-a\right)^2 ,$$

Table 5.5: *Comparison of mean squared prediction errors and mean absolute errors for income valuations according to WertV and multiple technique.*

Panel A: Mean Squared Prediction Errors (MSPE)			
	Variance	Bias	MSPE
WertV price to income value	0.129	0.135	0.147
θ_{PM} price to income value	0.167	0	0.167
WertV income value to price	0.079	-0.042	0.080
θ_{MP} income value to price	0.126	0	0.126
Panel B: Mean Absolute Prediction Errors (MAPE) in percent			
	MAPE	Percentage within 15%	
WertV price to income value	25.72%	44.64%	
θ_{PM} price to income value	29.56%	31.94%	
WertV income value to price	21.17%	45.35%	
θ_{MP} income value to price	27.80%	32.46%	

Note: Calculated for all 3835 objects with information on gross rents.

which measures the expected squared distance between the realizations of the predictor X and the predictive target a. It is composed of two terms: the variance of the predictor $\mathcal{V}[X]$ and the squared bias of the predictor. Applied to our implementation, the price to appraisal ratios—respectively the appraisal to price ratios—are the predictors and $a = 1$ is the predictive target.

Panel A in Table 5.5 reports the MSPEs for both income valuation techniques. Although the appraisals calculated with the multiple technique are unbiased, they are ranked in inferiorly, because valuations according to WertV have lower MSPEs. Panel B of Table 5.5 gives the mean absolute prediction errors (MAPE) for both valuation techniques, where the absolute prediction error is $|X - 1|$. In both cases, valuations according to WertV deliver smaller MAPEs and larger fractions of prediction errors that are at most 15%. These errors are of comparable size to the errors calculated by Kaplan and Ruback (1995, Table II) for meticulously prepared discounted cash flow valuations of companies.

Depending on their loss function, investors are not only interested in the first and second moments of the appraisal errors but in the entire distribution of these errors. We estimate the densities with kernel smoothing techniques

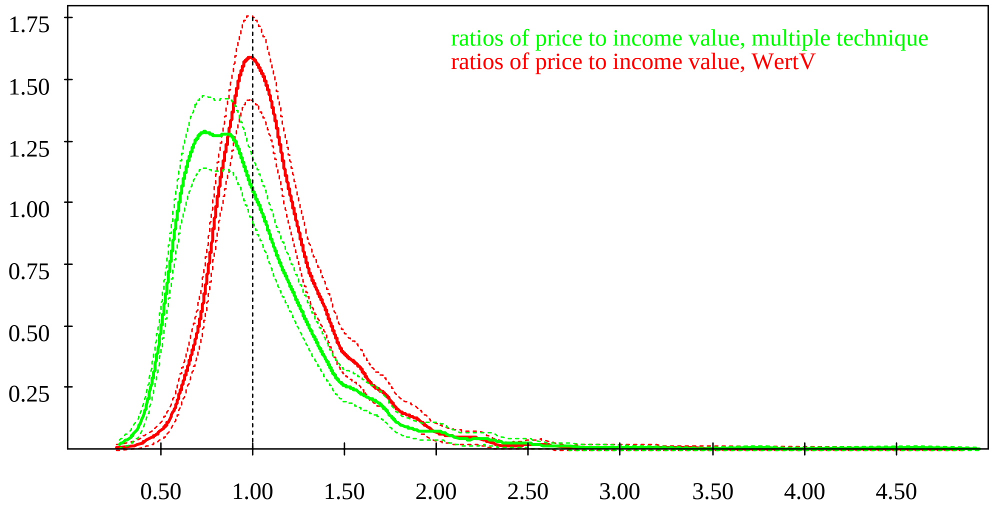

Figure 5.2: Nonparametric density estimates for ratios of price to income value according to WertV and for ratios of price to income value according to the multiple technique. Uniform confidence bands are at the 95% level.

(Härdle; 1991; Wand and Jones; 1995; Simonoff; 1996). Due to the fact that the market component M_t will be dependent over time, our observations are not independent. Hart and Vieu (1990) have shown that the least-squares cross validation criterion gives the asymptotically optimal bandwidth for dependent observations. However, as recommended by Müller (2000, p. 181), bandwidths were also determined by other selection methods. This is done with XploRe's `denbwsel` Quantlet. Comparing the resulting density estimates have revealed no differences with respect to the bandwidth selected by the least-squares cross validation criterion.

Figure 5.2 shows kernel density estimates with uniform confidence bands, see Müller (2000), for ratios of price to income value (WertV) and for ratios of price to income value (multiple technique). We use XploRe's `denxcb` Quantlet. Whereas the density of WertV ratios peaks close to 1—the modus is at about 0.974—the most likely realizations of multiple ratios are definitely smaller than 1. About 60% of multiple ratios are below 1, compared with 40% of WertV ratios. Multiple ratios are more 'bullish' than WertV ratios, because more income values overstate than understate the price. We conclude that it depends on the objectives of the investor which distribution of appraisal errors is more 'tolerable'. A ranking of the distribution for ratios

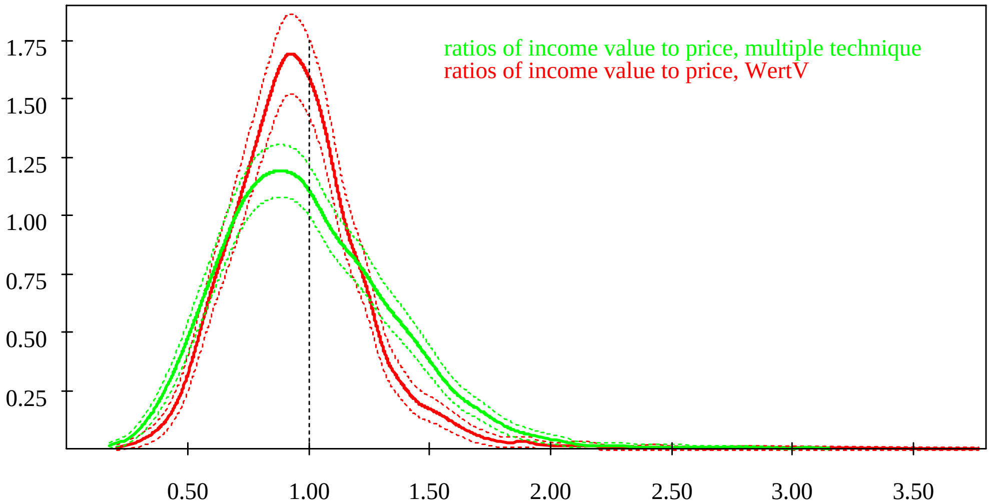

Figure 5.3: Nonparametric density estimates for ratios of income value to price according to WertV and for ratios of income value to price according to the multiple technique. Uniform confidence bands are at the 95% level.

of income value to prices is more obvious. Figure 5.3 shows kernel density estimates for ratios of income value to price. Once again, income values calculated with the multiple technique are more 'bullish' than income values calculated according to WertV. Whereas for the latter about 60% of the ratios lie below 1 (income values are smaller than prices), only 55% of the former lie below 1. It is obvious that large deviations from the true price are more likely for appraisals calculated with the multiplier technique. The tails of the corresponding density dominate the tails of the density of the WertV appraisals. If investors dislike the occurrence of large appraisal errors, they prefer valuation according to WertV.

Can we conclude that valuation according to WertV is definitely better? The multiple technique is an easy rule of thumb that needs little information. It ignores the remaining time of usage, object-specific operating costs and conditions of the property. Valuation according to WertV is much more information-intensive and needs such information. More information should lead to better appraisals and our results are not surprising. It is possible that lower valuation costs of the multiple technique will outweigh the higher accuracy of valuation according to WertV. However, without knowledge of economic losses associated with appraisal errors, it is impossible to decide on this question. At least, graphical representations of error distributions might be a valuable device for investors to decide about the preferable valuation technique.

5.5 General market conditions

We have shown that income valuation according to WertV is an accurate first step for ascertaining the market value. The second step consists in adjusting the income value with the general market conditions, where the appraiser should take several market indicators into account.

According to the present value model, the general market conditions are short-run deviations of the discount rates, which is represented in our two equation model (5.19)

$$\phi(L)m_t = \tilde{\kappa} + s_t\gamma + \xi_t$$

and

$$q_{n,t} = m_t + \varepsilon_{n,t} \; .$$

Both equations set up a state space model (SSM) (Harvey; 1989; Durbin and Koopman; 2001). Here, the first equation is the state equation and the second is the measurement equation. The characteristic structure of state space models relates a series of unobserved values to a set of observations. In our case, the state equation models the behavior of the general market conditions. The measurement equation relates the observed log ratios of all houses sold in period t to the behavior of the general market conditions. Figure 5.4 shows a nonparametric density estimate for the dependent variable in the SSM. When some parameters of the SSM are unknown, they can be estimated via maximum likelihood. Under the assumption of normality, the likelihood function—see equation (3.22)—can by evaluated with Kalman filter techniques (Harvey; 1989; Shumway and Stoffer; 1982, 2000). For our model, these parameters are the coefficients of the lag polynomial $\phi(L)$, the weights for the financial indicators γ, the constant $\tilde{\kappa}$, and the variances σ_ξ^2 and σ_ε^2.

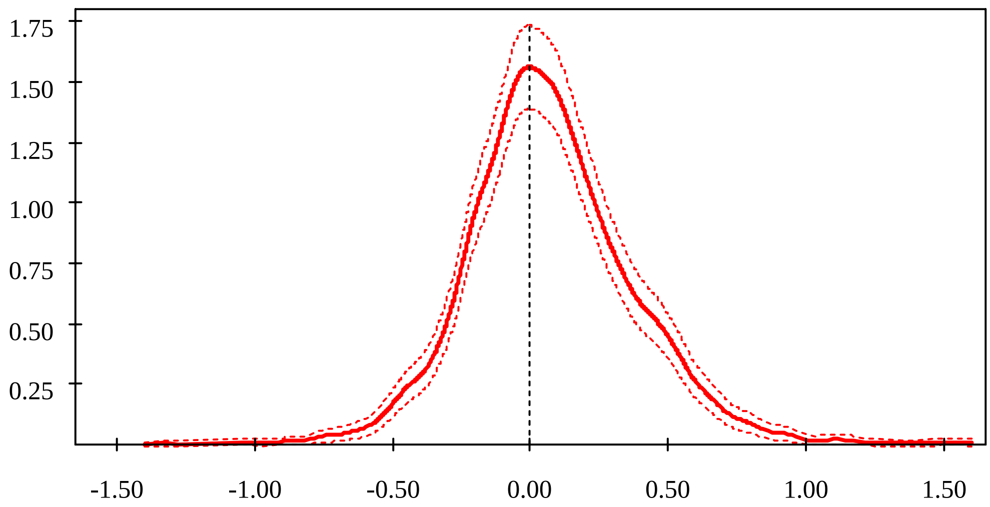

Figure 5.4: Nonparametric density estimate for log ratios of price to income value according to WertV. Uniform confidence bands are at the 95% level.

5.5.1 Market indicators

According to the Report on the Berlin real estate market of the GAA, the general market conditions are influenced by five-year mortgage rates, interest rates on credits, the consumer price index of the StaLa, building permissions and the number of transactions (Geschäftsstelle des Gutachterausschusses für Grundstückswerte in Berlin; 2001). The Report does not motivate the channels through which the market indicators influence the real estate market.

Most variables can be seen as proxies for economic risk that influences required returns of investments in residential real estate. Different to that, building permissions should have no influence on required returns, because anticipated changes in the stock of buildings should be incorporated in expected rents. To capture these anticipated changes we use the difference of the current growth rate of building permissions and its twelve-month average. The moving average can be interpreted as the 'normal' level of building permissions due to replacement and depreciation of older buildings. We use building permits for Berlin West to cover the sample period, which is no severe restriction because the largest part of our data comprises houses from that part. Interpreting the transaction volume as a proxy for incentives to buy due to newly announced tax brackets and subsidies, a higher volume should be accompanied by higher general market conditions.

To model the financing conditions on the market, we use the spread of the five-year mortgage rate and the capital market rate with the same maturity. These series are only obtainable since 1982:6. The spread can be interpreted as a risk premium for mortgage loans and thus for investments in real estate. Banks try to match the volume of mortgage credits by deposits with the same maturity. Given that interpretation, spreads should be stationary (Nautz and Wolters; 1996). Conducting an ADF test for the longer sample period, including a constant and no lags gives a test statistic of -4.58. Using MacKinnon's critical values, we can reject the hypothesis of a unit root at the 1% level (critical value is -3.46). The spread should have a positive influence on required returns and a negative relationship with the general market conditions. Real estate investors often use checking accounts for interim financing (Brauer; 1999). We use the interest rate for such credits with a withdrawal between 0.2 up to 1 Million German marks. Dividing by twelve and subtracting the expected monthly inflation rate gives roughly the real interest costs. The expected monthly inflation rate is given by the fitted values of an AR(12) model. Using seven lags and a constant, the ADF test statistic is -4.54. Using MacKinnon's critical values, we can reject the hypothesis of a unit root at the 1% level (critical value is -3.46). The effect of the real interest rate on required returns is not clear. Given tax deductability of interim interest payments, the effect of changes in real interest rates will be indeterminate.

Eventually, to control for information leads of the surveyors, we fit an AR(12) model for the inflation of the rent index and use the innovations—that are current values minus fitted values—as a measure of potential information leads of the surveyors. Given our results about the incompletely appraised age and the influence of the log size of the lot on q, we include both variables in the measurement equation. We do not control for the real gross rent, because that figure is seldom observed after 1995.

Fitting several specifications for the process of the general market conditions, comparing the value of the log likelihood function and the behavior of the state residuals, we choose the ARMA(1,1) specification. The log-likelihood function of the SSM, see 3.22, is evaluated with XploRe's `gkalfilter` Quantlet and is maximized with `nmBFGS`. Standard deviations are estimated via the Hessian matrix of the log-likelihood function evaluated at the estimates of the coefficients, see Hamilton (1994, 5.8). The Hessian is calculated numerically with XploRe's `nmhessian` Quantlet. Using a significance level of 1%, we cannot reject that the residuals behave like white noise and are normally distributed with a Jarque-Bera Statistic of 0.87 and a corresponding P-Value of 0.65.

Table 5.6: *Estimated SSM for the general market conditions for 1982:6 -2000:5.*

	Coefficient		t-Statistic	P-Value
$\hat{\phi}$	0.925		32.26	0.000
$\hat{\theta}$	-0.689		-6.07	0.000
$\widehat{\ln \sigma_\xi}$	-2.892		-20.56	0.000
$\widehat{\ln \sigma_\varepsilon}$	-1.265		-104.87	0.000
$\hat{\tilde{\kappa}}$	0.168		2.44	0.015
Log lot size	-0.015		-1.66	0.097
Age	-0.001		-2.67	0.008
Spread5	-3.953		-2.90	0.004
Real interest	5.759		1.68	0.092
Building permissions	-0.027		-1.55	0.121
Log number of transactions	0.015		2.52	0.012
Rent index	-0.986		-0.94	0.349
Diagnostics				
Log likelihood	2720.192	$\widehat{\sigma_\varepsilon^2}$		0.080
Observations	3629	$\widehat{\sigma_q^2}$		0.096

Notes: The market indicators are lagged by one month and demeaned. Building permissions is the deviation between the growth rate and its twelve-months moving average. Rent index gives the innovations of a fitted AR(12) model for the inflation rate of the rent index. Spread5 is the difference between mortgage and interest rate with 5-year maturity. Real interest is the difference between the monthly interest rate for check accounts and the expected monthly inflation rate.

Table 5.6 reports the results of the estimated SSM. The first coefficient is the effect of the lagged general market conditions on its current value and the second is the MA(1) coefficient. The disturbances account for about 83% in the variation of q. The remaining variation is due to the general market conditions and systematic appraising errors due to mis-adjusted remaining time of usage, where the negative effect due to miss-assessment of lot values is insignificant at the 5% level. In general, the result suggests that correct adjustment of general market conditions will increase the accuracy of ascertained market values.

As conjectured, the spreads between mortgage and interest rate have a depressing effect on the general market conditions. Interpreting the spread as risk premium, a higher risk premium increases required returns and has a negative effect on prices given the current expectations on net operating rents. The sign of the real interest rate coefficient is positive but insignificant at the 5% level. The positive sign might be puzzling because higher real rates make interim financing more expensive. Tax deductibility of interim interest payments may mitigate this effect and may make real estate investments more attractive compared with other investments. Building permissions have an insignificant coefficient. That is in accordance with the argument that changes in building permissions will influence the expected rents but not the required returns for investments in real estate. The number of transactions have a positive influence on the general market conditions. According to the experience of valuers from the GAA, they are a proxy for announced changes in tax privileges for investments in real estate. Eventually, the innovations in the rent index have an insignificant negative coefficient. So we reject that the surveyors of the GAA confound their backward-looking appraisals with current information.

Figure 5.5 shows the smoothed general market conditions. Up to 1989, prices and appraisals are in line, but afterwards they are out of touch for several years. That can be explained by political events. After the reunification of Germany in 1990 and the decision for Berlin as the Capital of the reunified Germany in 1991, some pundits expected that the city would scale to its position before the Second World War. Prices for real estate reached high levels, gross cap rates were down at 5% and required returns for investments in real estate were low. Income values according to WertV lagged behind because they are calculated with higher discount rates. However, income values catched up to prices relatively quickly when the outlook for real estate investments went gloomier and required returns reverse to their normal level.

5.6 Conclusion

Income valuation according to WertV is the most prominent valuation technique for income generating properties in Germany. Its motivation is the

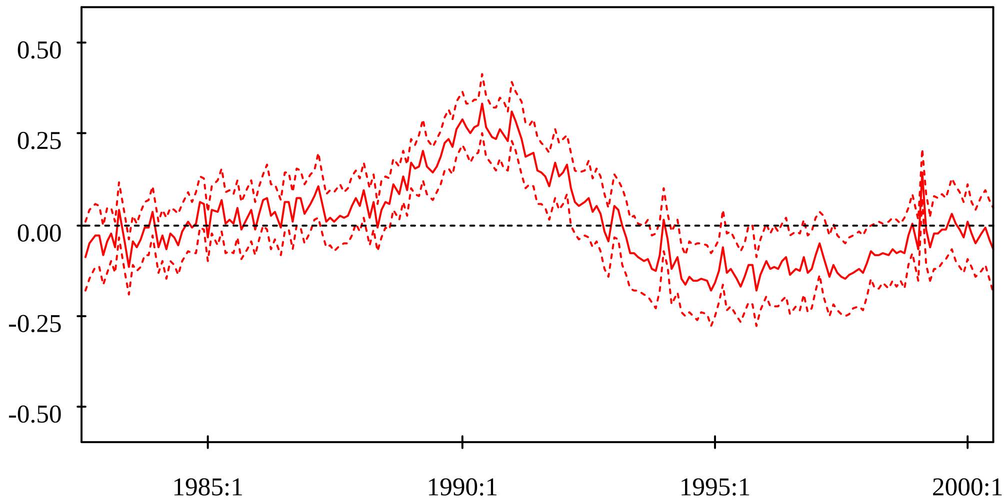

Figure 5.5: Smoothed general market conditions 1982:6-2000:5. Confidence bands are at the 95% level.

present value and the appraisal technique is codified in detail.

On average appraisals are 13% lower than transaction prices. Mis-assessed property specific characteristics explain in part deviations between prices and valuations. Compared with the simpler multiple technique, income valuation according to WertV seems preferable. The latter—although biased—gives smaller MSPEs and smaller MAPEs. Because the assessment of valuation accuracy depends on the investor specific loss function, we have estimated the densities of price valuation ratios. They can serve as a tool for investors to decide about their preferred valuation technique.

Income values are only the first step in ascertaining the market value of a property. Systematic deviation of prices and income values occur if short-run required returns deviate from their long-run rate. Valuers adjust for this in a second step. We have shown that systematic deviations explain about 17% of the total variance in price valuation ratios and are influenced by market indicators. These are good reasons to conclude that appraisals can be improved by adjusting for general market conditions.

The remaining 83% of variation are explained in part by disturbances due to unusual circumstances during the business dealing. Such circumstances occur if buyer and seller are affiliated or one party has a special interest in the property. Although we have included the transaction volume as an indicator for changes in taxation of investments in real estate, taxation—which is neglected in valuations according to WertV—may explain why income values understate prices on average. Given that tax advantages are capitalized into prices, the latter will be higher than they would be without such advantages. Although this seems a promising strategy to explore price appraisal deviations further, tax advantages are often investor-specific and it is impossible to tackle such effects without information on buyers and sellers.

5.7 Appendix

We want to inspect the partial effects of $D_{n,t}$, $L_{n,t}$, $\tau_{n,t}$, and θ_t on the income value (5.4). Is is obvious that $E_{n,t}$ increases in the net operating rent $D_{n,t}$ and in the value of the lot $L_{n,t}$.

Furthermore, we obtain

$$\frac{\partial E_{n,t}}{\partial \tau_{n,t}} = \left(\frac{1}{1+\theta_t}\right)^{\tau_{n,t}} \ln\left(\frac{1}{1+\theta_t}\right)\left(L_{n,t} - \frac{D_{n,t}}{\theta_t}\right) .$$

This expression is positive—with $\theta_t > 0$—if the last term is negative. § 20 WertV states that a surveyor has to set $E_{n,t} = L_{n,t}$ if $D_{n,t} \leqslant \theta_t L_{n,t}$ happens. That explains why all income values in our data set are positive.

If $E_{n,t} = L_{n,t}$, changing the remaining time of usage does not influence the income value at all. On the other hand, when $D_{n,t} > \theta_t L_{n,t}$, the remaining time of usage increases the income value. For easier interpretation, we assume that $D_{n,t} > \theta_t L_{n,t}$ is fulfilled for all data and that a larger $\tau_{n,t}$ would have yielded a larger income value. Moreover, we obtain

$$\frac{\partial E_{n,t}}{\partial \theta_t} = -\left\{1 - \left(\frac{1}{1+\theta_t}\right)^{\tau_{n,t}}\right\}\frac{D_{n,t}}{\theta_t} - \tau_{n,t}\left(\frac{D_{n,t}}{\theta_t} + L_{n,t}\right)\left(\frac{1}{1+\theta_t}\right)^{\tau_{n,t}} < 0 .$$

Chapter 6

An Empirical Analysis of Cost Valuation in Germany

6.1 Introduction

> "The incentive for new building can be measured by comparing the value of old homes with the cost of building new ones. The new ones won't be duplicates of the old, but will be close functional substitutes. We could expect residential investment to be sensitive to the housing q."
>
> James Tobin (1978, p. 425)

In Chapter 2, we mentioned that generally any of the three major valuation approaches, i.e. sales comparison, income, and cost approach, can be used for valuation of single-family houses. For example, our online prediction service MD*Immo is based on a combination of the sales comparison and the income approach. However, the cost approach is standard for appraising single-family houses in Germany, see Table 2.1.

According to the cost approach, the prospected transaction price of a property is figured out by calculating its replacement costs and by adjusting the replacement costs with a market-dependent adjustment factor. The rationale for using the cost approach for valuation is based on a simple economic argument: Any informed buyer would not pay more for a property than what it would cost to buy the land and build the structure (Brueggeman and Fisher; 2001, p. 243). But this simple argument ignores that construction needs time, that land for building purposes might be scarce, and that building a house needs more effort than buying an existing one. Thus, it is reasonable that prices and costs are linked only loosely.

Given the possible divergence of market prices for existent houses from the replacement costs, it is questionable if the cost approach is a convenient device for real estate valuation. What is needed is an economic model which explains why prices and costs should obey an equilibrating process.

What justifies that prices and replacement cost obey an equilibrating process and what influences short run deviations? Tobin's Q theory of investment delivers answers to these questions. Q is the ratio of prices for existing asset to its replacement costs, like prices for existing houses and construction costs. According to Tobin's theory, Q determines net investment and is an important figure for the transmission process between the valued stock of assets and the flow of new investment (Brainard and Tobin; 1968; Tobin; 1969; Tobin and Brainard; 1977). In equilibrium, prices must be equal to costs and $Q = 1$. Out of equilibrium, profit opportunities exist—do not exist—which induce net investments above—below—the equilibrium level. Tobin's theory stands in the Keynesian tradition and states that rigid prices propel an adjustment process of asset stock, which subsequently affects asset prices. However, as Hayashi (1982) has shown, Q can also be justified in a neoclassical setting with optimal capital accumulation where firms face adjustment cost. In the housing market, say, such adjustment costs exist because land for building purpose might be rare.

The Q theory gives an explanation why prices and replacement costs are closely linked. However, the empirical results in ascertaining factors that influence housing supply are mixed, see DiPasquale (1999). Modelling the demand for structures and neglecting the costs for land, Poterba (1984) and Topel and Rosen (1988) both found that construction cost do not play a role for investment. In both studies investment is significantly influenced by real house prices, which are prices for constant quality houses divided by the personal consumption deflator. Mayer and Somerville (2000) estimate a model that includes costs for land, but they still obtain that construction costs have no significant influence on investment. Moreover, all cited studies found significant coefficients for variables where no economic explanations are available. One may object that all of the above studies work with aggregated, nation-wide data and that real estate markets are regional. In a cross-sectional study for 39 US cities, Poterba (1991) found evidence that a 1% rise in construction costs leads to a increase of real house prices by the same amount.

To summarize, there is evidence that a relationship between prices and new construction exists, but replacement costs may play no role in this relationship. This is a severe objection against the cost approach for valuation, because it is based on the assumption that prices and replacement costs are closely related. If no relationship exists, should the cost approach be used for valuation?

In this chapter, we evaluate the accuracy of the cost approach with 6062 data from Berlin during 1995 to 2002. In Germany, valuation according to the cost approach is codified by law, so that all assessed values are derived with the same technique. We check for the accuracy of the appraised replacement costs. Deviation between prices and replacement costs should not depend on idiosyncratic factors like the size or the age of the building. All deviations should be systematic, given the current state of the market. Section 6.2 presents the economic model. Section 6.3 presents the data in detail and reports the estimation results. The final Section 6.4 concludes.

6.2 Cost approach

Valuation is the process of ascertaining the market value $V_{n,t}$ of the subject property n in period t. The market value, defined in the German Building Law (BauGB), is the price one should expect for a property given its characteristics and given usual business dealings (§ 194 BauGB). Let $P_{n,t}$ denote the price of property n in period t and let $\mathcal{E}_t[\cdot]$ denote the conditional expectation operator, we have

$$V_{n,t} = \mathcal{E}_t[P_{n,t}] \; .$$

Given that unusual circumstance may happen during the sale, the transaction price is

$$P_{n,t} = V_{n,t} U_{n,t} \; , \tag{6.1}$$

with $\mathcal{E}_t[U_{n,t}] = 1$. Here, $U_{n,t}$ captures unsystematic effects which may occur if one of the contracting parties has a special interest in obtaining the object or if there are personal relationships between seller and buyer. Valuers—according to personal communication—often use proportional discounts or surcharges, which justifies multiplicative disturbances, see also § 14, 25 of the German Regulation on Valuation (WertV). For a detailed discussion see Section 2.1.

According to WertV, replacement costs $C_{n,t}$ for property n in period t are the sum of the costs for the building $B_{n,t}$ and the price for the lot $L_{n,t}$

$$C_{n,t} = B_{n,t} + L_{n,t} \ .$$

$B_{n,t}$ is assessed after inferring the physical characteristics of the subject property, like its type and its size. Construction costs indices are used for inferring the current replacement costs. The replacement costs have to be adjusted for depreciation and the edificial condition of the structure. Techniques for adjusting the age of the building and its conditions are described in detail in the Guidelines on Valuation (WertR 91). Moreover, $B_{n,t}$ includes additional expenses like credit and notary fees, see § 22 WertV and Gottschalk (1999, A III). The value of the lot is assessed with the sales comparison approach by using prices of recently—comparable—sites. All costs that occur for the lot acquisition—like fees for real estate agents—are not scheduled.

After ascertaining the replacement costs, the appraiser is free to adjust $C_{n,t}$ to derive the ascertained market value by a market adjustment factor M_t, which considers the current situation on the real estate market (§ 3 Abs. 3 WertV). Let $V^a_{n,t}$ denote the ascertained market value, we obtain

$$V^a_{n,t} = M_t C_{n,t} \ .$$

The ascertained market value is thus related to replacement costs, but it is not necessarily equal to the costs.

We have derived in Section 2.2 that prices and costs are related through a function in net investments, see especially Result 2.5. This function was derived by inverting the net investment function as function of the ratio of price to costs. It is based on Tobin's Q theory, which delivers a motivation for the adjustment factor M_t. Q is defined as

$$Q_{n,t} \stackrel{\text{def}}{=} \frac{P_{n,t}}{C_{n,t}} \ , \tag{6.2}$$

where we plausibly assume that the costs are always positive. Given that prices are higher than cost, i.e, $Q > 1$, builders have profit opportunities by building new properties and offering it at market prices. The new investment increases the stock of existing properties and prices for it will decrease. Moreover, the additional demand for new land, construction material and workers may increase the cost for constructing new properties. Both will bring prices and costs together. If prices are below costs, i.e., $Q < 1$, builders have no incentive to offer new properties. Given depreciation, the stock of existing houses will decrease over time and prices for it will increase. Moreover, land, construction material and workers are not demanded, which may decrease the construction costs. Both will bring prices and costs together. In equilibrium $Q = 1$ and the rate of net investment will be equal to the natural growth rate of the economy (Tobin; 1978), see also appendix 6.5. Out of equilibrium, prices and costs are related by the factor M_t, which is a function of net investment, see Result 2.5. $M_t > 1$ if current net investment is higher than equilibrium investment and $M_t < 1$ if it is lower. If the market is in equilibrium, then $M_t = 1$.

Using the price equation (6.1), the definition of Q (6.2), and assuming that the ascertained market value equals the market value, one obtains

$$Q_{n,t} = M_t U_{n,t} \ . \tag{6.3}$$

In the short-run, we have

$$\mathcal{E}_t[Q_{n,t}] = M_t$$

and in the long-run

$$\mathcal{E}[Q_{n,t}] = 1 \ .$$

Observed ratios of prices and assessed replacement costs for individual properties can be used to test if the average ratio is one. The replacement cost should take into account all property-specific circumstances which will influence the replacement costs. These are: the age, the condition and the type of the building and the value of the lot.

Transaction prices may be influenced by unusual circumstances during the negotiations or the property may be rented. We have seen in Chapter 3 that such circumstances, which have nothing to do with the physical structure of the property, can influence transaction prices of single-family houses. Bequest and other personal motives lead to transaction prices where the inherent market value of a property plays only a secondary role. In these cases, prices are only noisy signals of market values. Corgel (1997) proposes to estimate market values with the hedonic approach. Therefore, in the empirical investigation, we will work additionally with ratios $Q^e_{n,t}$, where prices are replaced by estimated market values.

6.3 Empirical investigation

6.3.1 Data

The primary data set is provided by the Surveyor Commission for Real Estate in Berlin (GAA) and contains information on 6062 sales of single-family houses that occurred between January 1995 and July 2002 in Berlin. The data are stored in the non-public part of MD*Base, www.mdtech.de.

Table 6.1: *Summary statistics for transacted single-family houses in Berlin between 1995:1 to 2002:3.*

Panel A: Continuous Characteristics, Prices, and Appraisals						
	Mean	Median	Std. Dev.	Min	Max	Units
Lot size	555.4	500.0	334.2	100.0	7045.0	*Square metres*
Floor space	143.9	135.0	50.7	39.0	778.0	*Square metres*
Age	39.4	37	26.7	0	200	*Years*
Price	265.3	237.2	144.1	47.7	2454.2	*Thsd. EUR*
Cost value	300.4	268.5	162.5	49.6	1804.2	*Thsd. EUR*
Lot value	156.6	131.3	117.7	15.9	1706.1	*Thsd. EUR*
Building value	143.8	127.3	100.9	0	1102.1	*Thsd. EUR*

Panel B: Location and House Quality			
Simple location	30.7%	Average location	48.5%
Good location	20.0%	Excellent location	0.8%
Waterside	0.9%	Bad condition	9.6%
Normal condition	59.7%	Good condition	30.7%
Panel C: Availability at the Date of Sale			
Seller-occupied	81.8%	Rented out	1.4%
Purchase by former tenant	4.2%	New object	12.6%
Panel D: House Types			
Detached house	52.0%	Row house	26.2%
Observations	6062	Semi-detached house	21.8%

Notes: Original currency units for sales before 2002 are German marks which are converted to EUR by dividing with 1.95583. New object comprises properties which are sold in the year of construction or are still under construction in the year of the sale.

All observations have information on the transaction price, the size of the lot and the floor space, the age of the building and its condition, and the quality of the location. Personal circumstances are recorded in the data, like affiliation between seller and buyer or partition of an estate. Sales are observed for all but the inner-city districts Mitte and Friedrichshain-Kreuzberg, where only few single-family houses exist. The appraised cost value consists of the sum of the construction costs for the building in its current state plus the value of the lot. Because the construction price index that is used for appraising the construction costs has a quarterly frequency, we chose quarters as the time period of our analysis. This gives us at least 5, at most 329 and on average 195.6 observations per quarter.

Table 6.1 reports summary statistics for the data. Panel A presents information on lot size, floor space, age, and reports information on prices and appraised values. Since 2002 all monetary variables are recorded in EUR and we use it as common currency unit. Panel B reports location characteristics which were introduced in 1996 as prescribed variables into the GAA data base (AKS). They indicate the quality of the neighborhood for the respective property. Properties with an excellent location are all from Charlottenburg-Wilmersdorf and Steglitz-Zehlendorf. After Berlin's district reform in 2001 (Statistisches Landesamt Berlin; 2001, p. 8), these new districts comprise the former four South-West districts, see Figure 3.1. Most of the properties which are located on the waterside are from the districts Treptow-Köpenick and Spandau. Panel C reports the availability of a property for the buyer at the date of sale and Panel D gives information on the house types.

We have two quarterly construction costs indices, which are used for appraising replacement costs of a building. The first is a general construction cost index for dwellings in Berlin and the second is a special index for construction costs of single-family houses. Both series are provided by Berlin's Statistical Office (StaLa) in its Statistical Report M I 4.

6.3.2 Empirical results

Given that prices may be noisy signals for market values, we run a hedonic regression on prices and characteristics. Here, the transformations of the continuous variables lot size, floor space, and age are determined according to the procedure described in Section 3.3. The optimal values of the transformation parameters λ, see (3.12), are 0.5 (size of the lot), 0.5 (size of the floor space), and -1 (age of the building), respectively. The value of the R^2-like, standardized cross-validation criterion in (3.14) for these transformations is 0.713.

Table 6.2: *OLS estimates of optimal specification of hedonic regression for single-family houses 1995:1 to 2002:3.*

	Coefficient	t-Statistic	P-Value
$T_{0.5}$(lot size)	0.195	19.19	0.000
$T_{0.5}$(floor space)	0.231	29.52	0.000
T_{-1}(age)	-0.382	-19.83	0.000
Detached house	0.040	3.71	0.000
Average location	0.057	6.97	0.000
Good location	0.197	17.00	0.000
Excellent location	0.642	13.46	0.000
Waterside	0.296	6.59	0.000
Good condition	0.098	10.23	0.000
Bad condition	-0.249	-17.27	0.000
Rented out	-0.168	-4.30	0.000
Purchased by former tenant	-0.092	-4.31	0.000
Buyer legal entity	0.100	2.15	0.031
Seller public body	-0.138	-4.34	0.000
Seller building society	-0.111	-6.96	0.000
Seller legal entity	-0.032	-2.14	0.032
Personal circumstances	-0.160	-5.06	0.000
Charlottenburg-Wilmersdorf	0.261	4.59	0.000
Lichtenberg-Hohenschönhausen	-0.252	-12.03	0.000
Marzahn-Hellersdorf	-0.196	-13.31	0.000
Neukölln	0.098	7.70	0.000
Pankow	-0.200	-12.75	0.000
Reinickendorf	0.091	8.47	0.000
Steglitz-Zehlendorf	0.331	21.53	0.000
Tempelhof-Schöneberg	0.114	9.90	0.000
Treptow-Köpenick	-0.108	-6.54	0.000
	Diagnostics		
R^2	0.720	$\bar{R}^2$	0.717
F-Statistic	275.030	P-Value(F-Stat.)	0.000
Observations	6062	$\widehat{\sigma_\varepsilon^2}$	0.054

Notes: Dependent variable is the log price. $T_\lambda(\cdot)$ is the transformation function given in equation (3.12). t-Statistics are calculated with robust standard errors. Included overall constant and time dummies are not reported.

Table 6.2 reports the results from the hedonic regression, where the quarterly common price component is modelled by time dummies. The price elasticities for the continuous variables—evaluated at the respective sample means reported in Table 6.1—are 0.3% for lot size, 0.5% for floor space, and −0.1% for age. Most of the variables have reasonable signs. Simple location is the excluded location category and the coefficients reveal that all other location categories command premiums. Waterside is an amenity which increases the price. Normal condition is the excluded category for the variable which rates the edificial condition of the property. Rented properties, purchases by former tenants, and personal circumstances all command price rebates. The positive coefficient for the category 'buyer is a legal entity' seems puzzling because developers fall into this category. Normally, developers buy free sites and not existent single-family houses. They will only do so if it offers interesting restructuring or enlargement opportunities. This may explain the premium. Public bodies sell their properties with a rebate, which might indicate other than commercial interests. Building societies and developers often sell properties which are still under construction. This is a possible explanation for the estimated rebates. Eventually, the excluded district is Spandau and the districts dummies indicated the premiums—respectively rebates—of the other districts. Using the estimated common price components, i.e., the time dummies from the hedonic regression, a price index for the single-family houses in our data is constructed. The index is normalized to 100 for the first quarter in 1995. The index values

$$\hat{I}_t = \exp\left(b_t - 0.5\widehat{\sigma_t^2}\right)$$

are corrected for small-sample bias, where b_t is the estimated time dummy for quarter t and $\widehat{\sigma_t^2}$ is its estimated variance. The standard deviations for the index values are calculated with the delta method as the square of the first derivative of index values with respect to b_t times the variance of b_t, see Kennedy (1998, p. 37).

Figure 6.1 presents the price index in its first exhibit. The large jump from the first quarter to the second quarter in 1995 is explainable with the fact that we observe only five—low price—sales in 1995:1. For subsequent quarters, we have much more observations. The second exhibit in Figure 6.1 shows the general construction cost index for dwellings and the special index for single-family houses. It is obvious that both prices and construction costs have an downward trend.

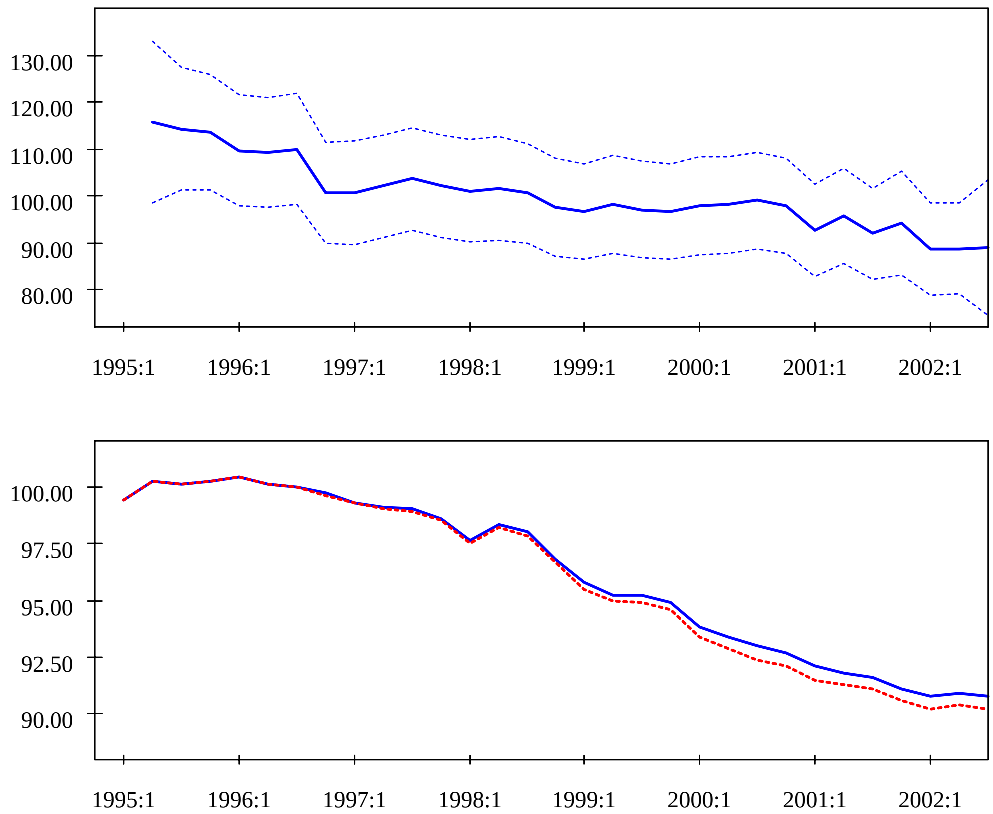

Figure 6.1: Quarterly single-family house prices and construction costs. First exhibit shows constant quality hedonic price index for Berlin single-family houses from 1995:2 to 2002:3. Confidence intervals at the 95% are calculated with the delta method. Second exhibit shows the construction price index for dwellings (solid) and single-family houses (dashed) from 1995:1 to 2002:3.

Table 6.3: *Summary statistics for ratios of price to cost value and for ratios of estimated market value to cost value. Cost values are appraised according to WertV.*

Panel A: Ratios of Price to Cost Value				
Mean	Standard deviation	Minimum	Median	Maximum
0.955	0.332	0.135	0.888	3.383
10% Quantile	90% Quantile	Skewness	Kurtosis	Number of obs.
0.624	1.367	1.679	7.773	6062
Panel B: Ratios of Estimated Market Value to Cost Value				
Mean	Standard deviation	Minimum	Median	Maximum
0.969	0.311	0.271	0.908	3.760
10% Quantile	90% Quantile	Skewness	Kurtosis	Observations
0.652	1.358	1.614	7.423	6062

Given the estimated coefficients from the hedonic regression, the market values can be estimated. For doing this, log market values are predicted by setting all dummies to zero which control for personal circumstances, rented objects, purchases by former tenants, and non-private buyers or sellers. All other characteristics remain unchanged and are evaluated with their respective implicit price. The log market values are transformed to market values. Eventually, the market values were divided by the respective cost value to derive $Q^e_{n,t}$.

Table 6.3 reports summary statistics for ratios of price to cost value and for ratios of estimated market value to cost value. In both cases, the sample means are below one. Calculating the t-Statistics for the hypothesis that expected Q ratios are one, we obtain figure of -10.517 (price to cost value) and -7.795 (estimated market value to cost value). In both cases, we reject the hypothesis that appraised cost values are unbiased predictors for prices and market values.

Table 6.3 shows that the figures for Qs calculated with estimated market values are not really different from the figures calculated with prices. The standard deviation of the former is expectedly smaller, because the estimated market values were adjusted for unusual circumstances. In most cases, unusual circumstances lower prices. Whereas 4073 Q ratios are smaller than one, only 3884 Q^e ratios are smaller than one. Figure 6.2 shows nonparametric density estimates for both ratios. As recommended by Müller (2000,

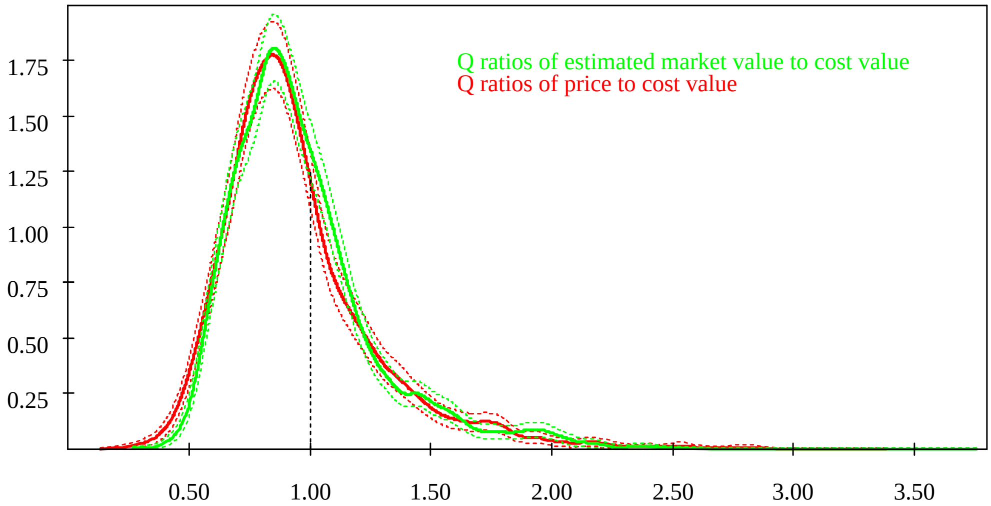

Figure 6.2: Nonparametric density estimates for ratios of price to cost value and for ratios of estimated market value to cost value. Uniform confidence bands are at the 95% level.

p. 181), bandwidths are determined by different selection methods and the resulting density estimates are nearly identical. This is done with XploRe's `denbwsel` Quantlet. The density estimates in Figure 6.2 reveal that both ratios are distributed similar.

Figure 6.3 shows average ratios of Q. It is obvious that both ratios behave very similar. Such Q's can reveal information about the current state of the market for single-family houses. According to the figure, prices are lower than replacement costs for the periods before the year 2000 and are higher thereafter. Such figures might be used as an indicator of the future development of net investment, see Corgel (1997). However, it is important that the cost values are assessed correctly. Otherwise average Qs are useless, because it will be confounded by mis-appraisement.

Are there property-specific characteristics which explain why transaction prices and appraised replacement costs diverge? According to equation (6.3), Q ratios are equal to the common market component M_t multiplied with the disturbance term which captures unusual circumstances during the business dealings. It is the very idea of the cost approach that all characteristics of the subject property n have to be assessed in the cost value $C_{n,t}$. This value

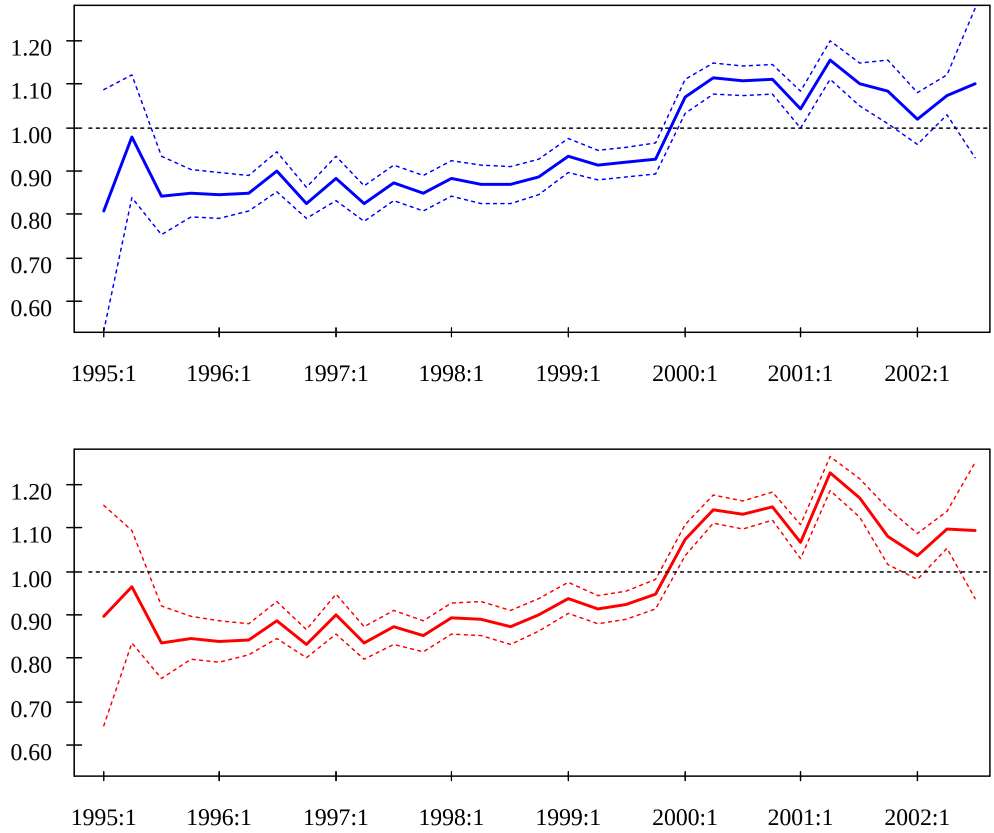

Figure 6.3: Quarterly average ratios of price to cost value and estimated market value to cost value. Confidence intervals are at the 95% level.

considers the location—via the assessed value of the lot—and the conditions of the building, like its age or its type. So, characteristics of the property should not explain the Q ratios. Taking logs on both sides of (6.3) gives

$$q_{n,t} = m_t + \varepsilon_{n,t} \ ,$$

where $q_{n,t} \stackrel{\text{def}}{=} \ln Q_{n,t}$, $m_t \stackrel{\text{def}}{=} \kappa + \ln M_t$, and $\varepsilon_{n,t} \sim (0, \sigma_\varepsilon^2)$. The constant κ guarentees that the disturbance term ε has an expected value of zero.

Running a regression for $q_{n,t}$ on property characteristics reveals that the characteristics have explanatory power. Table 6.4 reports the results. Some of the included variables control for unusual circumstances during the business dealings which lead to deviations between prices and market values.

Table 6.4: *OLS estimates of optimal regression specification for q ratios 1995:1 to 2002:3*

	Coefficient	t-Statistic	P-Value
T_{-1}(lot size)	-2.284	-24.27	0.000
T_2(floor space)	0.007	5.35	0.000
T_1(age)	0.184	34.36	0.000
Detached house	0.022	2.47	0.013
Row house	0.032	3.39	0.001
Average location	0.042	5.68	0.000
Good location	0.136	12.93	0.000
Waterside	0.259	6.27	0.000
Good condition	0.099	10.83	0.000
Bad condition	-0.094	-7.60	0.000
Rented out	-0.165	-4.69	0.000
Seller public body	-0.070	-2.48	0.013
Personal circumstances	-0.149	-4.68	0.000
Unusual transactions	-0.070	-2.57	0.010
Lichtenberg-Hohenschönhausen	-0.075	-4.57	0.000
Marzahn-Hellersdorf	0.068	4.84	0.000
Pankow	0.034	2.72	0.007
Reinickendorf	-0.040	-4.94	0.000
Steglitz-Zehlendorf	-0.066	-5.28	0.000
Treptow-Köpenick	0.082	5.99	0.000
	Diagnostics		
R^2	0.467	$\bar{R}^2$	0.462
F-Statistic	105.143	P-Value(F-Stat.)	0.000
Observations	6062	$\widehat{\sigma_\varepsilon^2}$	0.054

Notes: Dependent variable is $q = \ln Q$. $T_\lambda(\cdot)$ is the transformation function given in equation (3.12). t-Statistics are calculated with robust standard errors. Included overall constant and time dummies are not reported.

Comparing Table 6.4 with the hedonic regression reported in Table 6.2 reveals that they are of comparable magnitude. Personal circumstances depress prices by about 15% and rented properties change hands with a rebate of about 16%, which is in accordance with the results from the hedonic regression. The variable for unusual transactions was not significant in the hedonic regression. It controls for personal contributions of the buyer and several—quite heterogeneous—other unusual circumstances. Some of the variables which control for seller's and buyer's type are not significant at the 5% level in the above regression. All but one of these insignificant variables have the same sign as in the hedonic regression. Only the sign for the variable 'seller is a building society' changes. We see that the appraised cost values diminish some effects of unusual circumstances on prices.

More important are the other significant variables that control for the characteristics of the subject property. Given that the rationale of the cost approach is the correct market model and that the cost values are assessed correctly, none of these variables should have an influence on the q ratios. Without quarterly dummies, the R^2 for a regression with all of the above property characteristics is still high with 0.369.

Table 6.4 shows that the q ratios decrease with the size of the lot. Properties with large lots have smaller qs than properties with small lots. So, appraised lot values may not reflect the correct lot values. In addition to inaccuracies regarding the size of a lot, there are also inaccuracies regarding its location. Whereas an excellent location has no influence on the ratios, average and good location and waterside still have. All of these location characteristics, which have positive influences on transaction prices—see Table 6.2—are not assessed to the correct degree in the cost values, i.e. in the assessed lot value. The district dummies reveal mis-assessment of the location as well. Interestingly, all districts in the East part of Berlin have significant dummies, which indicates that it is difficult to figure out lot values for these districts. Comparable transaction data for the districts in the East part are not collected before 1990, the year of the German Reunification. This lack of recorded comparable transactions may explain why assessed lot values are inaccurate.

In addition to assessed lot values, values of buildings seem also to be inaccurately assessed. The type of the subject property has an influence on q, where the ratios are higher for detached and row houses. qs increase for buildings with large floor spaces and increase for older buildings. Moreover, the edificial conditions of the subject property have an influence on q. Bad conditions still command a rebate of about 10% compared to the cost value and properties in good conditions still command a premium of about 10%. The state of repair and the age should be completely assessed in the cost value, but our results suggests that this is not true.

Recall our discussion from Section 1.1 on the measurement of aggregated real estate: according to DiPasquale and Wheaton (1996, p. 4), assessed replacement costs of structures "may bear little resemblance to the actual market value of real estate", because it is very difficult to find correct figures for depreciation. "In addition, it is difficult to estimate the value of the land on which those structures are built because most observed transactions provide a single purchase price for the existing building and the land, with no breakdown by the land and structure components." Our results confirm this conjecture. Given our large set of individual data, cost values are inaccurate for market values, because they assess characteristics different to the market.

6.4 Conclusion

The cost approach for valuation has some drawbacks. First, it is difficult to figure out replacement costs for existent buildings, because construction costs for new buildings have to be adjusted for depreciation. Second, prices for existing properties and the costs for building new ones can diverge for several periods. It is a difficult task to adjust for such divergence. German appraisers often use inverse average Q ratios for adjusting indicated value. The adjustment factors are provided by local Surveyor Commissions and are calculated with historical transaction data, for Berlin see Senatsverwaltung für Bauen, Wohnen und Verkehr (1999). But it is obvious that such backward-looking adjustment factors may deliver inaccurate ascertain market values. Third, and most important, it is questionable if existing properties and new ones are close functional substitutes. That is a severe problem for the validity of the cost approach.

Whereas the income and the sales comparison approach value directly the characteristics of the subject property, the cost approach values the market value only indirectly. But given taxes and assistance, buying an existent property and building a new one will be no close substitutes. In Germany, sales of existent houses are taxed by the tax on purchase of real estate, but the building costs are exempted from this tax. Only the price of the lot is taxed. Sales of existing houses and newly built houses are promoted differently, where the promotion of a newly built house is higher, see the Ownership Promotion Law (§ 9 EigZulG). Given different treatment of existent houses and newly built ones, it seems questionable if the long-run Q will be equal to one. It is plausible that the favorable status of newly built properties will lead to a long-run Q which is smaller than one. If this is true, appraisals with the cost approach have to be adjusted with a factor which considers this fact. Currently, such an adjustment is not prescribed in the German WertV.

The empirical investigation has shown that cost values are only inaccurate predictors of prices. Cost values are on average higher than prices. More important, property characteristics still explain deviations between prices and cost values to a large degree.

6.5 Appendix

In the literature on the real option theory of investment one can find another version of Q. Investment is not triggered until this Q is well above one, see Dixit and Pindyck (1994) and Capozza and Li (2002). We can relate this version of Q to our application: according to the real option theory of investment (McDonald and Siegel; 1986; Pindyck; 1991), land is an option on project development and L is the value of waiting for favorable investment conditions (Quigg; 1993). L can be expressed as a function of the price of the developed property. Using this, one can derive a Q which relates the price of a project to structure costs only.

Such a Q makes sense for applications where option values are not observable. But L is observable. It can be included directly in the costs C. If this is done, the critical Q that triggers investment is again one, see Dixit and Pindyck (1994, 2.C).

Bibliography

Bach, S. and Bartholmai, B. (1998). Immobilienvermögen privater Haushalte in Deutschland 1995, *DIW Wochenbericht* **65**(35): 630–642.

Bailey, M. J., Muth, R. F. and Nourse, H. O. (1963). A regression method for real estate price index construction, *Journal of the American Statistical Association* **58**: 933–942.

Berens, W. and Hoffjan, A. (1995). Wertermittlung von Immobilien auf Basis vollständiger Finanzpläne, *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* **47**: 373–395.

Berger, T., Englund, P., Hendershott, P. H. and Turner, B. (2000). The capitalization of interest subsidies: Evidence from Sweden, *Journal of Money, Credit, and Banking* **32**: 199–217.

Berndt, E. R. (1991). *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley, Reading M.A.

Berndt, E. R., Griliches, Z. and Rappaport, N. J. (1995). Econometric estimates of price indexes for personal computers in the 1990's, *Journal of Econometrics* **68**: 243–268.

Brainard, W. C. and Tobin, J. (1968). Pitfalls in financial model building, *American Economic Review Papers & Proceedings* **58**: 99–122.

Brauer, K.-U. (1999). Immobilienfinanzierung, *in* K.-U. Brauer (ed.), *Grundlagen der Immobilienwirtschaft: Recht, Steuern, Marketing, Finanzierung, Bestandsmanagement, Projektentwicklung*, Gabler, Wiesbaden, pp. 347–409.

Brealey, R. A. and Myers, S. C. (2000). *Principles of Corporate Finance*, sixth edn, McGraw-Hill, New York.

Brown, G. R. and Matysiak, G. A. (2000). *Real Estate Investment. A Capital Market Approach*, Pearson Education, Harlow.

Bruce, D. and Holtz-Eakin, D. (1999). Fundamental tax reform and residential housing, *Journal of Housing Economics* **8**: 249–271.

Brueggeman, W. B. and Fisher, J. D. (2001). *Real Estate Finance and Investments*, eleventh edn, McGraw-Hill Irwin, New York.

Bulwien (2001). Immobilienindex 1975 bis 2001, Auszug, Bulwien AG, München. www.bulwien.de.

Bunke, O., Droge, B. and Polzehl, J. (1999). Model selection, transformations and variance estimation in nonlinear regression, *Statistics* **33**: 197–240.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Capozza, D. R. and Li, Y. (2002). Optimal land development decisions, *Journal of Urban Economics* **51**: 123–142.

Case, B. and Quigley, J. M. (1991). The dynamics of real estate prices, *Review of Economics and Statistics* **73**: 50–58.

Case, B. and Szymanoski, E. J. (1995). Precision in house price indices: Findings of a comparative study of house price index methods, *Journal of Housing Research* **6**: 483–496.

Case, K. E., Quigley, J. M. and Shiller, R. J. (2001). Comparing wealth effects: The stock market versus the housing market, *Working Paper W8606*, National Bureau of Economic Research.

Case, K. E. and Shiller, R. J. (1987). Prices of single-family homes since 1970: New indexes for four cities, *New England Economic Review* (September/October): 45–56.

Case, K. E. and Shiller, R. J. (1989). The efficiency of the market for single-family homes, *American Economic Review* **79**: 125–137.

Case, K. E., Shiller, R. J. and Weiss, A. N. (1993). Index-based futures and options markets in real estate, *Journal of Portfolio Management* **19**: 83–92.

Chinloy, P., Cho, M. and Megbolugbe, I. F. (1997). Appraisals, transaction incentives, and smoothing, *Journal of Real Estate Finance and Economics* **14**: 89–111.

Clapp, J. M. and Giaccotto, C. (2002). Evaluating house price forecasts, *Journal of Real Estate Research* **24**: 1–26.

Cochrane, J. H. (2001). *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Corgel, J. B. (1997). Real estate Q, *Working paper*, School of Hotel Administration, Cornell University.

Corgel, J. B., Ling, D. C. and Smith, H. C. (2001). *Real Estate Perspectives. An Introduction to Real Estate*, fourth edn, McGraw-Hill Irwin, New York.

Cropper, M. L., Deck, L. B. and McConnell, K. E. (1988). On the choice of functional form for hedonic price functions, *Review of Economics and Statistics* **70**: 668–675.

Crosby, N. (2000). Valuation accuracy, variation and bias in the context of standards and expectations, *Journal of Property Investment & Finance* **18**: 130–161.

Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society – Series B* **39**: 1–38.

Deutsche Börse (2002). Leitfaden zu den Aktienindizes der Deutschen Börse, Gruppe Deutsche Börse Frankfurt am Main. Version 4.6.

Deutsche Bundesbank (1998). Entwicklung und Struktur des gesamtwirtschaftlichen Kapitalstocks, *Deutsche Bundesbank Monatsbericht* (November): 27–39.

Deutsche Bundesbank (1999). Zur Entwicklung der privaten Vermögenssituation seit Beginn der neunziger Jahre, *Deutsche Bundesbank Monatsbericht* (Januar): 33–50.

Deutsche Bundesbank (2002). Der Wohnungsmarkt in den neunziger Jahren, *Deutsche Bundesbank Monatsbericht* (Januar): 29–39.

DID (2002). DIX Deutscher Immobilien Index 2002, Deutsche Immobilien Datenbank GmbH, Wiesbaden. www.dix.de.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business & Economic Statistics* **13**: 253–263.

Diewert, W. E. (1976). Exact and superlative indexes, *Journal of Econometrics* **4**: 115–145.

Diewert, W. E. (1998). Index number issues in the consumer price index, *Journal of Economic Perspectives* **12**(1): 47–58.

DiPasquale, D. (1999). Why don't we know more about housing supply?, *Journal of Real Estate Finance and Economics* **18**: 9–23.

DiPasquale, D. and Wheaton, W. C. (1996). *Urban Economics and Real Estate Markets*, Prentice Hall, Englewood Cliffs, New Jersey.

Dixit, A. K. and Pindyck, R. S. (1994). *Investment under Uncertainty*, Princeton University Press, Princeton, New Jersey.

Dombrow, J., Knight, J. and Sirmans, C. (1997). Aggregation bias in repeat-sales indices, *Journal of Real Estate Finance and Economics* **14**: 75–88.

Dotzour, M. G. (1988). Quantifying estimation bias in residential appraisal, *Journal of Real Estate Research* **3**(3): 1–11.

Dubofsky, D. A. (1992). *Options and Financial Futures. Valuation and Uses*, McGraw Hill, New York.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*, Oxford Statistical Science Series 24, Oxford University Press, Oxford.

Engle, R. F., Lilien, D. M. and Watson, M. (1985). A DYMIMIC model of housing price determination, *Journal of Econometrics* **28**: 307–326.

Engle, R. F. and Watson, M. W. (1981). A one-factor multivariate time series model of metropolitan wage rates, *Journal of the American Statistical Association* **76**: 774–781.

Englund, P., Hwang, M. and Quigley, J. M. (2002). Hedging housing risk, *Journal of Real Estate Finance and Economics* **24**: 167–200.

Englund, P. and Ioannides, Y. M. (1997). House price dynamics: An international empirical perspective, *Journal of Housing Economics* **6**: 119–136.

Englund, P., Quigley, J. M. and Redfearn, C. L. (1998). Improved price indexes for real estate: Measuring the course of swedish housing prices, *Journal of Urban Economics* **44**: 171–196.

Fisher, I. (1922). *The Making of Index Numbers*, Houghton-Mifflin, Boston.

Geschäftsstelle des Gutachterausschusses für Grundstückswerte in Berlin (2001). *Bericht über den Berliner Grundstücksmarkt 2000*, Kulturbuch-Verlag, Berlin.

Goetzmann, W. N. (1992). The accuracy of real estate indices: Repeat sale estimators, *Journal of Real Estate Finance and Economics* **5**: 5–53.

Gottschalk, G.-J. (1999). *Immobilienwertermittlung*, C.H. Beck, München.

Gourieroux, C. and Jasiak, J. (2001). *Financial Econometrics. Problems, Models, and Methods*, Princeton University Press, Princeton, New Jersey.

Gourieroux, C. and Monfort, A. (1997). *Time Series and Dynamic Models*, Cambridge University Press, Cambridge.

Graff, R. A. and Young, M. S. (1999). The magnitude of random appraisal error in commercial real estate valuation, *Journal of Real Estate Research* **17**: 33–54.

Gravelle, H. and Rees, R. (1992). *Microeconomics*, second edn, Longman, London and New York.

Greene, W. H. (2000). *Econometric Analysis*, fourth edn, Prentice Hall, Upper Saddle River, New Jersey.

Griliches, Z. (ed.) (1971). *Prices Indexes and Quality Change: Studies in New Methods of Measurement*, Harvard University Press, Cambridge, MA.

Halvorsen, R. and Pollakowski, H. O. (1981). Choice of functional form for hedonic price equations, *Journal of Urban Economics* **10**: 37–49.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Härdle, W. (1991). *Smoothing Techniques. With Implementation in S*, Springer Series in Statistics, Springer, Berlin, New York.

Härdle, W., Klinke, S. and Müller, M. (eds) (2000). *XploRe – Learning Guide*, Springer, Berlin, New York.

Hart, J. D. and Vieu, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data, *Annals of Statistics* **18**: 873–890.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation, *Econometrica* **50**: 213–224.

Hendershott, P. H. and White, M. (2000). The rise and fall of housing's favored investment status, *Journal of Housing Research* **11**: 257–275.

Herman, S. W. (2001). Fixed assets and consumer durable goods for 1925–2000, *Survey of Current Business* (September): 27–38.

Hill, R. C., Knight, J. and Sirmans, C. (1997). Estimating capital asset price indexes, *Review of Economics and Statistics* **79**: 226–233.

Homburg, S. (1994). Staatseingriffe auf dem Wohnungsmarkt und Marktunvollkommenheiten. Miete versus Selbstnutzung, *in* Expertenkommission Wohnungspolitik (ed.), *Wohnungspolitik auf dem Prüfstand. Materialband zum Gutachten im Auftrag der Bundesregierung*, Bonn.

Institut für Städtebau, Wohnungswirtschaft und Bausparwesen (2001). DEIX Deutscher Eigentums-Immobilien-Index, Institut für Städtebau, Wohnungswirtschaft und Bausparwesen Berlin. www.ifs-staedtebauinstitut.de.

Isakson, H. R. (2002). The linear algebra of the sales comparison approach, *Journal of Real Estate Research* **24**: 117–128.

Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions 1*, Wiley, New York.

Johnston, J. (1984). *Econometric Methods*, third edn, McGraw-Hill, New York.

Kaplan, S. N. and Ruback, R. S. (1995). The valuation of cash flow forecasts: An empirical analysis, *Journal of Finance* **50**: 1059–1093.

Kennedy, P. (1998). *A Guide to Econometrics*, fourth edn, Blackwell, Oxford.

Konüs, A. A. (1939). The problem of the true cost of living, *Econometrica* **7**: 10–29.

Koopman, S. J., Shephard, N. and Doornik, J. A. (1999). Statistical algorithms for models in state space using SsfPack 2.2., *Econometrics Journal* **2**: 107–160.

Lancaster, K. J. (1966). A new approach to consumer theory, *Journal of Political Economy* **74**: 132–157.

Ling, D. C. and Naranjo, A. (1997). Economic risk factors and commercial real estate returns, *Journal of Real Estate Finance and Economics* **15**: 283–307.

Lütkepohl, H. (1996). *Handbook of Matrices*, Wiley, Chichester.

MacKinnon, J. G. (1991). Critical values for cointegration tests, *in* R. F. Engle and C. W. J. Granger (eds), *Long-Run Economic Relationships. Readings in Cointegration*, Advanced Texts in Econometrics, Oxford University Press, Oxford, pp. 267–276.

Malpezzi, S. (2002). Hedonic pricing models and house price indexes: A selective and applied review, *in* A. O'Sullivan and K. Gibb (eds), *Housing Economics and Public Policy*, Blackwell, Oxford. Forthcoming.

Mayer, C. J. and Somerville, C. T. (2000). Residential construction: Using the urban growth model to estimate housing supply, *Journal of Urban Economics* **48**: 85–109.

McDonald, R. and Siegel, D. (1986). The value of waiting to invest, *Quaterly Journal of Economics* **101**: 707–728.

McParland, C., Adair, A. and McGreal, S. (2002). Valuation standards: A comparison of four European countries, *Journal of Property Investment & Finance* **20**: 127–141.

Meese, R. A. and Wallace, N. E. (1997). The construction of residential housing price indices: A comparision of repeat-sales, hedonic-regression, and hybrid approaches, *Journal of Real Estate Finance and Economics* **14**: 51–73.

Merlo, A. and Ortalo-Magné, F. (2002). Bargaining over residential real estate: Evidence from England, *Working paper*, Department of Economics University of Pennsylvania.

Miles, D. (1994). *Housing, Financial Markets and the Wider Economy*, Series in Financial Economics and Quantitative Analysis, Wiley, New York.

Moryson, M. (1998). *Testing for Random Walk Coefficients in Regression and State Space Models*, Contributions to Statistics, Physica-Verlag, Heidelberg.

Muellbauer, J. (1974). Household production theory, quality, and the "hedonic technique", *American Economic Review* **64**: 977–994.

Müller, M. (2000). Smoothing methods, *in* W. Härdle, S. Klinke and M. Müller (eds), *XploRe – Learning Guide*, Springer, Berlin, New York, pp. 169–204.

Nautz, D. and Wolters, J. (1996). Die Entwicklung langfristiger Kreditzinssätze: Eine empirische Analyse, *Kredit und Kapital* **29**: 481–510.

O'Sullivan, A. (2000). *Urban Economics*, fourth edn, McGraw-Hill, New York.

Palmquist, R. B. (1980). Alternative techniques for developing real estate price indexes, *Review of Economics and Statistics* **62**: 442–448.

Pindyck, R. S. (1991). Irreversibility, uncertainty, and investment, *Journal of Economic Literatur* **29**: 1110–1148.

Pollakowski, H. O. (1995). Data sources for measuring house price changes, *Journal of Housing Research* **6**: 377–387.

Poterba, J. M. (1984). Tax subsidies to owner-occupied housing: An asset-market approach, *Quarterly Journal of Economics* **99**: 729–752.

Poterba, J. M. (1991). House price dynamics: The role of tax policy and demography, *Brookings Papers on Economic Activity* **1991**(2): 143–183.

Quigg, L. (1993). Empirical testing of real option-pricing models, *Journal of Finance* **48**: 621–640.

RICS (2002). Property valuation: The Carsberg report, Royal Institution of Chartered Surveyors. www.rics.org.uk.

Ring Deutscher Makler (1989-1999). RDM Preisspiegel, Ring Deutscher Makler Bundesverband e. V., Hamburg. various issues, www.rdm.de.

Romer, D. (1996). *Advanced Macroeconomics*, McGraw-Hill Advanced Series in Economics, McGraw-Hill, New York.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy* **82**: 34–55.

Rüchardt, K. (2001). *Der Beleihungswert. Bedeutung, Anforderungen, Ermittlung und Verwendung von Beleihungswerten für Immobilien*, Schriftenreihe des Verbandes deutscher Hypothekenbanken Band 13, Fritz Knapp, Frankfurt am Main.

Schulz, R. and Werwatz, A. (2002). A simple state space model of house prices, *in* W. Härdle, T. Kleinow and G. Stahl (eds), *Applied Quantitative Finance. Theory and Computational Tools*, Springer, Berlin, New York, pp. 283–307.

Schwann, G. M. (1998). A real estate price index for thin markets, *Journal of Real Estate Finance and Economics* **16**: 269–287.

Searle, S. R. (1971). *Linear Models*, Wiley, New York.

Senatsverwaltung für Bauen, Wohnen und Verkehr (1999). Faktoren zur Anpassung des Sachwertes von Grundstücken mit Eigenheimen an die Lage auf dem Grundststücksmarkt, *Amtsblatt für Berlin* **49**(65): 5004–5011.

Senatsverwaltung für Stadtentwicklung (2000). Für die Wertermittlung erforderliche Daten, *Amtsblatt für Berlin* **50**(15): 1064–1068.

Senatsverwaltung für Stadtentwicklung und Investitionsbank Berlin (2002). *Der Berliner Wohnungsmarkt. Entwicklungen und Strukturen 1991–2000*, Regioverlag, Berlin.

Sheffrin, S. M. (1996). *Rational Expectations*, Cambridge Surveys of Economic Literature, second edn, Cambridge University Press, Cambridge.

Sheppard, S. (1999). Hedonic analysis of housing markets, *in* P. Cheshire and E. S. Mills (eds), *Handbook of Regional and Urban Economics, Volume 3: Applied Urban Economics*, Handbooks in Economics 7, North-Holland, Amsterdam, pp. 1595–1635.

Shiller, R. J. (1993a). *Macro Markets. Creating Institutions for Managing Society's Largest Economic Risks*, Clarendon Press, Oxford.

Shiller, R. J. (1993b). Measuring asset values for cash settlement in derivative markets: Hedonic repeated measures indices and perpetual futures, *Journal of Finance* **48**: 911–931.

Shiller, R. J. and Weiss, A. N. (1999). Evaluating real estate valuation systems, *Journal of Real Estate Finance and Economics* **18**: 147–161.

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis* **3**: 253–264.

Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*, Springer Texts in Statistics, Springer, Berlin, New York.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*, Springer Series in Statistics, Springer, Berlin, New York.

Spanos, A. (1999). *Probability Theory and Statistical Inference. Econometric Modeling with Observational Data*, Cambridge University Press, Cambridge.

Statistisches Landesamt Berlin (2001). *Statistisches Jahrbuch 2001*, Kulturbuch-Verlag, Berlin.

Summers, L. H. (1981). Inflation, the stock market, and owner-occupied housing, *American Economic Review Papers & Proceedings* **71**: 429–434.

Sydsæter, K., Strøm, A. and Berck, P. (2000). *Economists' Mathematical Manual*, third edn, Springer, Berlin.

Thomas, M. (1995). Immobilienwertbegriffe in Deutschland und Großbritannien, *Die Bank* (5): 263–268.

Tobin, J. (1969). A general equilibrium approach to monetary theory, *Journal of Money, Credit, and Banking* **1**: 15–29.

Tobin, J. (1978). Monetary policies and the economy: The transmission mechanism, *Southern Economic Journal* **44**: 421–431.

Tobin, J. and Brainard, W. C. (1977). Asset markets and the cost of capital, *in* R. Nelson and B. Balassa (eds), *Economic Progress: Private Values and Public Policy. Essays in Honor of William Fellner*, North-Holland, Amsterdam, pp. 235–262.

Topel, R. and Rosen, S. (1988). Housing investment in the united states, *Journal of Political Economy* **96**: 718–740.

Triplett, J. E. (1987). Hedonic functions and hedonic indexes, *in* J. Eatwell, M. Milgate and P. Newman (eds), *The New Palgrave: A Dictionary of Economics*, Vol. 2, Macmillan, London, pp. 630–634.

U.S. Bureau of Economic Analysis (1999). Fixed Reproductible Tangible Wealth in the United States, 1925–94, U.S. Government Printing Office, Washington DC.

Verband Deutscher Makler (2001). VDM Immobilienpreisspiegel 2001, Verband Deutscher Makler für Grundbesitz Hausverwaltung und Finanzierungen e.V. www.vdm.de.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, Monographs on Statistics and Applied Probability 60, Chapman & Hall, London.

Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models, *Journal of Econometrics* **23**: 385–400.

Webb, C. (1988). A probabilistic model for price levels in discontinuos markets, *in* W. Eichhorn (ed.), *Measurement in Economics. Theory and Applications of Economic Indices*, Physica-Verlag, Heidelberg, pp. 137–156.

Whitehead, C. M. E. (1999). Urban housing markets: Theory and policy, *in* P. Cheshire and E. S. Mills (eds), *Handbook of Regional and Urban Economics, Volume 3: Applied Urban Economics*, Handbooks in Economics 7, North-Holland, Amsterdam, pp. 1559–1594.

Wu, L. S.-Y., Pai, J. S. and Hosking, J. R. M. (1996). An algorithm for estimating parameters of state-space models, *Statistics & Probability Letters* **28**: 99–106.

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfaßt und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.

Rainer Schulz
3. Januar 2003