FACULTEIT ECONOMISCHE EN
TOEGEPASTE ECONOMISCHE
WETENSCHAPPEN

KATHOLIEKE
UNIVERSITEIT
LEUVEN

ADVANCED CLASSIFICATION AND TIME-SERIES METHODS IN MARKETING

Proefschrift voorgedragen tot
het behalen van de graad van
Doctor in de Toegepaste
Economische Wetenschappen

door

Aurélie LEMMENS

Nummer 221 Mei 2006

Doctoral Committee

Prof. Dr. Christophe Croux (advisor)
K.U. Leuven

Prof. Dr. Marnik G. Dekimpe (co-advisor)
K.U. Leuven and Tilburg University (The Netherlands)

Prof. Dr. Tammo H.A. Bijmolt
University of Groningen (The Netherlands)

Prof. Dr. Willy Gochet
K.U. Leuven

Prof. Dr. Sunil Gupta
Columbia University (USA)

Prof. Dr. Stefan Stremersch
Erasmus University Rotterdam (The Netherlands)

Acknowledgements

"The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.
Sir William Bragg (1862 - 1942)

The book you have in your hands would never have existed without the close collaboration and the strong support of several people I owe all my gratitude to.

First of all, I am mostly grateful to my advisor, Christophe Croux, who offered me the opportunity to enrol in the Ph.D. program at K.U.Leuven. In my opinion, I could not have dreamed of a better research environment in my home country. During the last four years, Christophe devoted many of his - yet precious - time to guide me through the process of getting my degree. He did not only have the patience to teach me a great deal of statistical methods, to show me how important it is to be statistically rigorous, but he also provided me with the tools to further learn and widen my knowledge of statistics by myself. The latter is the probably the most precious thing to become an independent researcher, ... even if I do still expect to come back to him with many new questions in the next future!

I also would like to express my thanks and deepest appreciation to Marnik Dekimpe who acted as a second advisor for me. His deep knowledge and dedication to the marketing research strongly impressed me and stimulated me to start an academic career in marketing myself. I guess that his passion must be highly contagious. The collaboration with Marnik was especially insightful to learn how a good piece of marketing research should be written, from the right positioning into the marketing literature to the writing of the managerial implications.

I am also deeply indebted to my other dissertation committee members, Tammo Bijmolt, Willy Gochet, Sunil Gupta and Stefan Stremersch. It has been a great honor for myself to have such a prestigious committee. I really appreciate the time you devoted to my dissertation, from the proof-reading of the chapters to your highly appreciated presence at my defense. Your insightful comments have provided excellent guidance and thoughtful reflection, not only with respect to my dissertation work, but with other research projects as well.

A special thanks also goes to Barbara Deleersnyder for her support and numerous advices. I especially remember her kind guidance during my first marketing conference in Glasgow. I also would like to take this opportunity to thank my colleagues at K.U.Leuven who make this a stimulating and friendly place to work. I would like to express my appreciation to Kristel Joossens who provided me with some useful linguistic and IT support. I also enjoyed the very nice collaboration with Sarah Gelper and the kindness of my office roommates, Jan Adem and Johan Van Kerckhoven. Many thanks also to my marketing colleagues, Kathleen Cleeren, Lien Lamey and Maarten Gijsenberg for the pleasant moments we spent together at conferences.

Furthermore, I gratefully acknowledge the financial support provided by the "Fonds voor Wetenschappelijk Onderzoek" and the Research Fund K.U.Leuven.

Last but not least, I am of course indebted to my family for encouraging me during my studies, and especially to Didier Gosset for his unconditional support, his interest in my work, the many proof-readings requested from him and, above all, his patient support regarding my professional choices wherever they are bringing us to.

A little and silly thank to Harry and Muchu
who kept my feet warm during the long
hours spent in front of my computer.

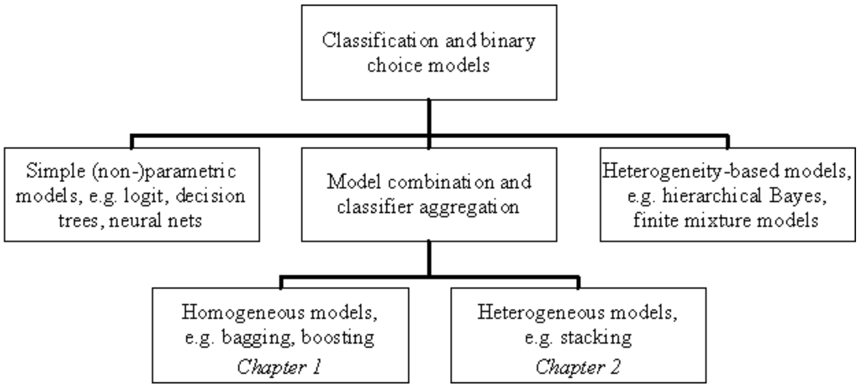
Aurélie Lemmens

Preface

This collection of essays investigates diverse marketing issues ranging from customer retention to the elaboration of pan-European strategies, through the use of advanced statistical modeling techniques. The first part focuses on classification methods while the second is grounded in the time-series framework.

Classification issues are common in marketing, spreading from consumer choice modeling to market segmentation. In Figure A, we depict the variety of classification models that marketers can apply to solve such issues. One of the most popular and extensively used is probably the logit model (see e.g. Andrews, Ainslie, and Currim, 2002). A few authors have also considered the use of nonparametric models in marketing, like decisions trees (Haughton and Oulabi, 1997) or neural networks (West, Brockett and Golden, 1997). Recently, marketers have identified a need for models that account for heterogeneity in consumer response (Allenby and Rossi, 1999). These include finite mixture models (Wedel and Kamakura, 2000) and hierarchical Bayes techniques (Yang and Allenby, 2003). In the first part of this manuscript, we focus on a third kind of classification method based on the principle of *classifier aggregation.* The idea is to combine a number of choice models that can be of the same nature (homogeneous models) or of different kinds (heterogeneous models). Both types of combination are consecutively reviewed in Chapters 1 and 2.

Figure A: *Classification issues and binary choice models in marketing*

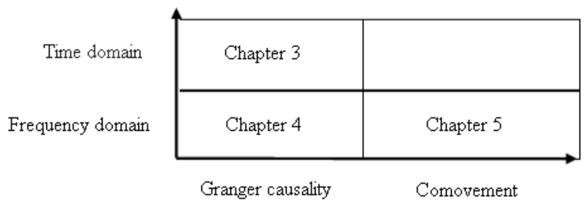


In Chapter 1, we investigate the use in marketing of bagging and boosting, two homogeneous model combination methods that conceptually originated in the statistical machine learning literature. We examine their performance in predicting customers' churn behavior for an anonymous U.S. wireless telecommunications company. The results indicate that bagging and boosting significantly improve accuracy in predicting churn. This higher predictive performance can ultimately be turned into incremental profits for companies using these methods. Although churn can lead dramatic financial losses, customer defection is still - statistically speaking - a rare event. In the first chapter, we also demonstrate that the use of a balanced sampling scheme is highly recommended in this context, but requires an appropriate bias correction.

In Chapter 2, we focus on the aggregation of classification models of different nature (e.g. decision tree, neural network, logit model). In particular, we define a *stacking* algorithm to find the optimal weighted average of different classifiers, yielding a systematically better cross-validated error rate on the calibration data. Furthermore, we compare this aggregation method with other stacking methods. Finally, we study the potential of bagging as a way to improve the performance of a stacked classification model. Stacking and bagging after stacking both lead to high predictive performance for a range of 12 different classification tasks (e.g. credit scoring, breast cancer detection, spam e-mailing detection). They have a strong potential to be successful also when applied to common marketing issues, like customer retention or consumer choice modeling.

In turn, the second part of this manuscript is based on time-series methods. As outlined in Figure B, the use of time-series analysis in marketing can be categorized along two dimensions. As far as the first dimension is concerned, we can make a distinction between studies grounded in the time or in the frequency domain. Most research in marketing has been located in the time domain, a few exceptions being e.g. Bronnenberg et al. (2004), and Deleersnyder et al. (2004). A distinction can also rely on whether the focus is on the *co-movement* between several time series, or on the direction of the relationship between the time-dependent variables. The latter closely relates to the concept of Granger causality (Granger, 1969). Figure B positions Chapters 3, 4 and 5 with respect to these two dimensions.

Figure B: *Time-series methods in marketing*



These chapters focus on the understanding of the mechanisms of interaction and influence within the European market place, using the Business and Consumer Tendency Surveys published by the European Commission. In particular, Chapters 3 and 4 are based on an industry-related indicator, the production expectations, and study its predictive content regarding actual production data. As it is intuitively obvious that actual production is partially predictable from previous expectations, the critical issue turns out to be whether or not this predictive quality holds over and above simple time-series extrapolation of past production data (Hanssens and Vanden Abeele, 1987). In other words, we measure the *incremental* forecasting contribution of the surveys, which are optional, expensive and time-consuming, using a Granger-causality framework. Whether or not the surveys have significant predictive content depends on both the respondents' efficient use of information and the extent to which they influence the agents' decisions. Both components are difficult to disentangle. In studying the Granger causality, we use monthly data, hereby reducing the temporal aggregation problem. Indeed, it is known (see e.g. Wei, 1982) that larger time scales lead to substantial problems in terms of lag structure and causality. In particular, temporal aggregation can change a one-sided causality into a two-sided feedback relationship. Using a monthly sampling frequency, we can circumvent this kind of problem.

Previous research has studied the predictive content of the expectation variables included in those business and consumer surveys through bivariate, within-country, Granger-causality tests. These tests have resulted in mixed conclusions. In Chapter 3, we extend previous research in various ways as we (i) explicitly allow for cross-country influences, and (ii) do so using both bivariate and multivariate Granger-causality tests. Specifically, the multivariate El Himdi - Roy (1997) test is adapted to jointly test the forecasting value of multiple Production Expectation series and to assess whether part of this joint effect is indeed due to cross-country influences. We also determine which countries' expectation series have most “clout” in predicting the production levels in the other member countries, or have highest “receptivity”, in that their production levels are Granger caused by the other countries' expectations. The results indicate strong evidence of Granger causality at the multi-country level. In addition, the analysis also reveals interesting cross-country relationships, which could be advantageous exploited in further research. The existence of cross-country predictive content can be explained by the increasing free transfer of ideas, products and technologies at the European level (Mahajan and Muller, 1994), resulting in both cross-country word-of-mouth effects (Tellis et al, 2003) and economic interdependence (as reflected in an intra-EU15 trade/GDP ratio of over 16%). Due to these factors, industrial optimism/pessimism in one country, as reflected in the evolution of its production expectations, may well affect the production accounts in another country (potentially via the accounts in the former country).

In Chapter 4, we gain further insights into the predictive content of the European production expectations in decomposing the Granger causality over the frequency band. To do so, we develop a bivariate spectral Granger-causality test that can be applied at each individual frequency of the spectrum. The spectral approach to Granger causality has the distinct advantage that it allows to disentangle different Granger-causality relationships over different frequencies. The results indicate that the predictive content of these surveys vary with the time horizon considered.

In contrast to the previous chapters, the last piece of research takes a closer look at the European consumer and focuses on the co-movement of the European Consumer Confidence Indicators. The ongoing European unification, along with recent advances in consumer mobility and communication technology, have suggested that the European Union can be treated as a single market to fully exploit the potential synergy effects from pan-European marketing strategies (Leeflang and van Raaij, 1995). However, some authors have also argued that European countries continue to differ considerably from each other, economically speaking but especially, as far as cultural identity is concerned (de Mooij and Hofstede, 2002).

Previous research, which mostly used domain-specific segmentation bases, has resulted in mixed conclusions. In Chapter 5, a more general (that is, independent from the domain in question) segmentation base is adopted as we consider the homogeneity in the European countries' Consumer Confidence Indicators. Since the segmentation basis is remote from the product context, the actionability of the segmentation is rather low. However, this general segmentation is intended to provide useful prior information to managers when knowledge from domain-specific segmentation studies related to the market into consideration is limited or even absent. In addition, rather than analyzing more traditional static similarity measures, we adopt the concepts of dynamic correlation and cohesion between countries. The short-run fluctuations in consumer confidence are found to be largely country specific. However, a myopic focus on these fluctuations may inspire management to adopt multi-country strategies, foregoing the potential longer-run benefits from more standardized marketing strategies. Indeed, the Consumer Confidence Indicators become much more homogeneous as the planning horizon is extended. However, this homogeneity is found to remain inversely related to the cultural, economic and geographic distances among the various member states. Hence, pan-regional rather pan-European strategies are called for.

This collection of essays gives rise to several directions for future research. As far as the first part is concerned, an area that is worthy of further attention is the development of the bagging and boosting techniques in the time-series or panel data context. Indeed, finite mixture models and hierarchical Bayes techniques have been repeatedly found to be very powerful in that context as they consider the heterogeneity in consumer response. It could be of interest to modify the bagging and boosting algorithms to disentangle individual effects from the random errors. Further research should also investigate how to construct classification rules that optimize marketing-related performance measures, like e.g. the top-decile lift. As this measure is related directly to profitability, it can be of high interest for marketers to use a classification model that specifically optimizes the top-decile lift, rather than e.g. the error rate.

Also the second part of this work offers future research perspectives. Given the European component of these studies, the question of the convergence of the European Union is of a particular interest. Extending Chapter 3, it would be straightforward (by splitting the data in two parts) and, hopefully, insightful to compare the degree of cross-country Granger causality before and after the introduction of the Euro. Given that Chapters 4 and 5 have provided further evidence that the relationship between two variables depends on the planning horizon, the question also remains whether and how these findings generalize to other contexts. In particular, it can be of interest to analyze cross-country influences and mixing behavior for new product diffusion over different planning horizons.

To conclude, I hope this work helps bridging the gap between the statistics and marketing research community and serves as a useful starting point for further research.

The various chapters of this manuscript can be found in:

Aurélie Lemmens and Christophe Croux (2006), Bagging and Boosting Classification Trees to Predict Churn, *Journal of Marketing Research*, 43(2).

Christophe Croux, Kristel Joossens and Aurélie Lemmens (2004), Bagging a Stacked Classifier, *Proceedings in Computational Statistics*, Ed. J. Antoch, Heidelberg, Physica-Verlag, 839-855.

Aurélie Lemmens, Christophe Croux and Marnik G. Dekimpe (2005), On the Predictive Content of Production Surveys: A Pan-European Study, *International Journal of Forecasting*, 21(2), 363-375.

Aurélie Lemmens, Christophe Croux and Marnik G. Dekimpe (2005), Decomposing Granger Causality over the Spectrum, *Working Paper*, submitted.

Aurélie Lemmens, Christophe Croux and Marnik G. Dekimpe (2005), The European Consumer: United in Diversity?, *Working Paper*, submitted.

References

[1] Allenby, G.M. and Rossi, P.E. (1999), Marketing Models of Heterogeneity, *Journal of Econometrics*, 89, 57-78.

[2] Andrews, R.L., Ainslie, A. and Currim, I.S. (2002), An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity, *Journal of Marketing Research*, 39, 479-87.

[3] Bronnenberg, B.J., Mela, C. and Boulding, W. (2004), The Periodicity of Competitor Pricing, *Working paper series at the Andersen School of Management*, UCLA.

[4] Deleersnyder, B., Dekimpe, M.G., Sarvary, M. and Parker, P.M. (2004), Weathering Tight Economic Times: The Sales Evolution of Consumer Durables over the Business Cycle, *Quantitative Marketing and Economics*, 2, 347-383.

[5] De Mooij, M. and Hofstede, G. (2002), Convergence and Divergences in Consumer Behavior: Implications for International Retailing, *Journal of Retailing*, 78, 61-69.

[6] El Himdi, K. and Roy, R. (1997), Tests for Non-Correlation of two Multivariate ARMA Time Series, *The Canadian Journal of Statistics*, 25, 233-256.

[7] Granger, C.W.J. (1969), Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, 37, 424-438.

[8] Hanssens, D.M. and Vanden Abeele, P.M. (1987), A Time-Series Study of the Formation and Predictive Performance of EEC Production Survey Expectations, *Journal of Business and Economic Statistics*, 5, 507-519.

[9] Haughton, D. and Oulabi, S. (1997), Direct Marketing Modeling with CART and CHAID, *Journal of Direct Marketing*, 11, 42-52.

[10] Leeflang, P.S.H. and van Raaij, F. (1995), The Changing Consumer in the European Union. A Meta-Analysis, *International Journal of Research in Marketing*, 12, 373-387.

[11] Mahajan, V. and Muller, E. (1994), Innovation Diffusion in a Borderless Global Market: Will the 1992 Unification of the European Community Accelerate Diffusion of New Ideas, Products, and Technologies?, *Technological Forecasting and Social Change*, 45, 221-235.

[12] Tellis, G.J., Stremersch, S. and Yin E. (2003), The International Takeoff of New Products: The Role of Economics, Culture and Country Innovativeness, *Marketing Science*, 22, 188-208.

[13] Wedel, M. and Kamakura, W.A. (2000), *Market Segmentation: Conceptual and Methodological Foundations*, 2d ed. Boston: Kluwer Academic Publishers.

[14] Wei, W.W.S. (1982), The Effects of Systematic Sampling and Temporal Aggregation on Causality. A Cautionary Note, *Journal of the American Statistical Association*, 77, 316-319.

[15] West, P.M., Brockett, P.L. and Golden, L.L. (1997), A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice, *Marketing Science*, 16, 370-91.

[16] Yang, S. and Allenby, G.M. (2003), Modeling Interdependent Consumer Preferences, *Journal of Marketing Research*, 40, 282-94.

Table of Contents

Part I

Advanced Classification Models and Marketing Issues

Chapter 1

Bagging and Boosting Classification Trees to Predict Churn

1.1 Introduction

Classification issues are common in marketing literature. One of the most frequent topics envisioned as a classification task is consumer choice modeling (see e.g. Chung and Rao 2004; Corstjens and Gautschi 1983; Currim, Meyer and Le 1988; Guadagni and Little 1983; Kalwani, Meyer and Morrison 1994). The current study considers a *binary* choice problem, namely the prediction of customer churn behavior.

Several classification models exist, but one of the most popular is the (binary) logit model, which has been used extensively in marketing to solve binary - or multiple - choice problems (see e.g. Andrews, Ainslie, and Currim 2002). More sophisticated models, which take into account the heterogeneity in consumer response, include finite mixture models (see e.g. Andrews and Currim 2002; Wedel and Kamakura 2000), and hierarchical Bayes techniques (see e.g. Arora, Allenby, and Ginter 1998; Yang and Allenby 2003). For binary choice problems, these approaches require the availability of panel data (i.e. data from several observations over time on multiple customers). However, in many applications (including the current one), a customer is observed only once over time, which makes it impossible to disentangle the individual effects from the random errors (Donkers et al. 2006).

In this chapter, we bring to the attention of marketers the *bagging* and *boosting* classification models that originated in the statistical machine-learning literature. Bagging (Breiman 1996) consists of sequentially estimating a binary choice model - called a *base classifier* in machine learning - from resampled versions of a given calibration sample. The obtained classifiers form a committee from which a final choice model can be derived by simple aggregation. Although bagging is simple and easy to use, more sophisticated variants also exist. *Stochastic gradient boosting* (Friedman 2002) is one of the latest developments thus far, and includes weights in the resampling procedure.

Although bagging and boosting have received increasing attention in various fields (e.g. Friedman, Hastie and Tibshirani 2000, for the UCI machine-learning archive; Nardiello, Sebastiani and Sperduti 2003, for text categorization; Varmuza, He and Fang 2003, for use in chemometrics; or Viane, Derrig and Dedene 2002, for an application in fraud claim detection), to the best of our knowledge, the marketing literature does not contain any reference (yet) to such models. Therefore, we attempt to fill this gap by empirically investigating whether bagging and stochastic gradient boosting can challenge more traditional choice models. In particular, we examine their performance in predicting customers' churn behavior for an anonymous U.S. wireless telecommunications company.[1] To evaluate the predictive accuracy of our churn model, we consider not only the misclassification rate - which may be misleading for rare events, such as churn - but also the *Gini coefficient* and *the top-decile lift.*

Churn is a marketing-related term that characterizes whether a current

[1]This database was provided by the Teradata Center for Customer Relationship Management at Duke University during the Duke/NCR Churn Modeling Tournament.

customer decides to take his or her business elsewhere (in the current context, to defect from one mobile service provider to another). As with many other sectors (e.g. the newspaper business), churn is an important issue for both the U.S. and the European wireless telecommunications industry. Monthly churn rates amount to approximately 2.6% (Hawley 2003), as a result of increased competition, lack of differentiation, and saturation of the market. Because the cost of replacement of a lost wireless customer amounts to \$300-\$700 (depending on the source of information, see e.g. Snel 2000) in terms of sales support, marketing, advertising, and commissions, churn may have damaging consequences for the financial wealth of companies. However, predicting churn enables the elaboration of targeted retention strategies to limit these losses (Bolton, Kannan and Bramlett 2000, Ganesh, Arnold and Reynolds 2000, Shaffer and Zhang 2002). For example, specific incentives may be offered to the most risky customers' segment (i.e. the most inclined to leave the company), with the hope that they remain loyal. Other scientific studies also point out the advantage of customer retention as a lower-cost operation than attracting new customers (Athanassopoulus 2000; Bhattacharya 1998; Colgate and Danaher 2000).

Despite the financial consequences that a 2% monthly churn rate may lead to, customer defection is still - statistically speaking - a rare event. Consequently, when the churn predictive model is estimated on a random sample of the customer population, the vast majority of nonchurners in this *proportional* calibration sample (i.e. the number of churners in this randomly drawn sample is *proportional* to the real-life churn proportion) dominates the statistical analysis, which may hinder the detection of churn drivers, and eventually decrease the predictive accuracy. To address this issue, the calibration sample size can be increased. However, this solution is usually not optimal (see the Results section; King and Zeng 2001a). A better solution to this issue consists in applying a selective sampling scheme to increase the number of churners in the calibration sample. Such a sampling scheme is called *balanced sampling* (or *stratified sampling* in King and Zeng 2001a, b). Theoretically, a potentially better-performing classifier could be obtained from such a sample, especially for small sample sizes (see e.g. Donkers, Franses and Verhoef 2003, King and Zeng 2001a,b). We investigate whether these findings are still valid for large sample sizes.

The estimation of a classification model from a balanced sample usually overestimates the number of churners in real life. Several methods exist to correct this bias (see e.g. Cosslett 1993; Donkers, Franses and Verhoef 2003; Franses and Paap 2001, pp.73-75; Imbens and Lancaster 1996; King and Zeng 2001a,b; Scott and Wild 1997). However, most of these corrections are dedicated to traditional classification methods, such as the binary logit model. Therefore, we subsequently discuss two easy correction methods for bagging and boosting, from which marketers can benefit to predict churn.

In summary, we investigate the following research questions: (i.a) Do the recent developments in statistical machine learning outperform the traditional binary logit model in predicting churn? If so, (i.b) what financial gains can be expected from this improvement? And (i.c) what are the more relevant churn drivers, or *triggers*, for which marketers can watch? Moreover, we propose (ii.a) two bias correction methods for balanced samples, and investigate (ii.b) how they comparatively perform. Finally, we investigate (iii) whether a choice model estimated on a balanced sample, with the bias appropriately corrected for, outperforms a choice model estimated on a proportional sample in large sample configurations.

The remainder of the chapter is organized as follows. The next section contains a description of the data. The three subsequent sections outline the bagging and boosting models, the bias correction methods for balanced sampling schemes, and the assessment criteria, respectively. We then empirically answer the aforementioned questions and offer some conclusions.

1.2 Data

The study is performed on a data set provided by the Teradata Center at Duke University. This database contains three data sets of mature subscribers (i.e. customers who had been with the company for at least six months) to a major U.S. wireless telecommunications carrier. The variable we attempt to predict is whether a subscriber churns during the period of 31-60 days after the sampling date (we know that the actual reported average monthly churn rate is approximately 1.8%). A delay of one month in measuring the churn variable is justified because the implementation of proactive customer retention incentives requires some time. In this case, marketers would have a one-month delay to target and retain customers before they churn. The churn response is coded as a dummy variable, where $y = 1$ if the customer churns and $y = -1$ if otherwise.

The two first data sets are used as calibration samples of 51,306 observations each.[2] The first data set is a *proportional calibration sample* (the proportion of churners in the sample is approximately 1.8%), and the second contains an *oversampled* number of churners such that the number of churners is perfectly *balanced* by the number of nonchurners. Selected at a future point in time, the third data set contains 100,462 customers, 1.8% of whom are churners. We use this third set as a validation (thus, we do not used it in our estimation) holdout sample to evaluate the performance of the prediction rules constructed from one of the aforementioned calibration samples. All samples contain a different set of customers.

To predict customers' churn potential, U.S. wireless operators usually take into account between 50 and 300 subscriber variables as explicative factors (Hawley 2003). From the high number of explicative variables contained in the initial database (171 variables), we retained 46 variables,

[2]Originally, the second data set contained 100,000 observations, but we reduced its size (by taking a random subset from it) to ensure a fair comparison between both calibration samples.

including 31 continuous and 15 categorical variables. The retained predictors include behavioral (e.g. the average monthly minutes of use over the previous three months, the total revenue of a customer account, the base cost of a calling plan), company interaction (e.g. mean unrounded minutes of customer-care calls), and customer demographic (e.g. the number of adults in the household, the education level of the customer) variables (for an overview, see Table 1.1). We selected the variables by first excluding all variables that contained more than 30% of missing values. Among the remaining variables, we selected those with the most potential relevance, following the results of a principal components analysis.[3] Note that for an equal comparison, we consider the exact same set of variables for all investigated models.

The handling of missing values is operated differently for the continuous and the categorical predictors. For the continuous variables, we imputed the missing values by the mean of the nonmissing ones. Because not answering a question may be as informative as a specific response, for each observation, we added an extra predictor that indicated whether there was at least one imputation. For categorical predictors, we created an extra level for each of them that indicated whether the value was missing.

1.3 The Bagging and Boosting Models

Both bagging and boosting originate from the machine-learning research community, and are based on the principle of *classifier aggregation*. This idea was inspired by Breiman (1996), who found gains in accuracy by combining several base classifiers, sequentially estimated from perturbed

[3]Because the purpose of this chapter is to investigate the comparative performance of different models, we do not provide further details about variable selection, which mainly serve to reduce computation time. Some experiments indicated that the performance of the classification rules barely changed, regardless of whether we implemented a variable selection procedure.

Table 1.1: *Description of the churn predictors.*

Behavioral Predictors	Company Interaction Predictors	Customer Demographics
Billing adjusted total revenue over the life of the customer (“total revenue over life”)	Having responded to an offer in the mail (yes/no)	Age of the first household member (“age”)
Mean number of attempted calls placed (“mean attempted calls”)	Mean minutes of use of customer care calls	Estimated income
Percentage change in monthly minutes of use versus previous three-month average (“change in monthly minutes of use”)	...	Social group
Mean total monthly recurring charge (“base cost of the calling plan”)		Marital status
Average monthly minutes of use over the previous six months (“average monthly minutes of use [six months]”)		Geographic area
Mean number of completed calls (“mean completed calls”)		Account spending limit
Mean number of peak calls (“mean peak calls”)		Children in the household (yes/no)
Total number of months in service (“months in service”)		Dwelling unit type
Mean number of inbound calls less than one minute (“mean inbound calls less one minute”)		Number of days of current equipment (“Equipment days”)
Mean of overage revenue (“mean overage revenue”)		Refurbished or new handset
Mean number of monthly minutes of use (“mean monthly minutes of use”)		Current handset price (“handset price”)
Mean unrounded minutes of use of outbound wireless to wireless calls (“mean monthly minutes wireless to wireless”)		...
...		

versions of the calibration sample. Among the several possible alternatives of base classifiers, classification trees (also known as CART, Breiman et al. 1984) are a sensible choice (Breiman 1996). Their use is not widespread in marketing literature (for exceptions, see Baines et al. 2003; Currim, Meyer and Le 1988; Haughton and Oulabi 1997), though they are powerful nonparametric methods. In recent years, statistical theory has been elaborated to provide a theoretical background for these techniques (e.g. for bagging, see Bühlmann and Yu 2002; for boosting, see Friedman, Hastie and Tibshirani 2000; for a comprehensive review, see Hastie, Tibshirani, and Friedman 2001).

For the sake of conciseness, the following subsection contains a brief description of the bagging algorithm. In the next subsection, we provide further details about the main differences between bagging and stochastic gradient boosting, one of the most sophisticated versions of boosting to date. An in-depth description of this method can be found in Friedman (2002).

1.3.1 Bagging

Bagging (i.e. a term derived from Bootstrap AGGregatING) is, by far, the simplest technique to upgrade, or to *boost*, the performance of a given choice model. We denote the calibration sample as $Z = \{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_N, y_N)\}$, where N is the number of customers in the calibration sample. In this expression, $x_i = (x_{i1}, \ldots, x_{ik}, \ldots, x_{iK})$ represents a vector that contains the K predictors for customer i, and y_i (equal to 1 or -1) indicates whether this customer i will churn. A base classifier $\hat{f}$ is estimated from this calibration sample, giving a score value of $\hat{f}(x)$ to each customer, where x are the characteristics of this subscriber. This score value indicates the risk to churn associated with each customer. For a specified cutoff value τ, we can predict customers as churners or nonchurners by computing

$$\hat{c}(x) = sign(\hat{f}(x) - \tau), \tag{1.1}$$

which takes values of +1 or -1. If $\hat{f}(x_i)$ is larger than τ, customer i is classified as a churner, while if $\hat{f}(x_i)$ is smaller than τ, customer i is predicted as a nonchurner. When we use a classification tree as base classifier, the score is given by $\hat{f}(x) = 2\hat{p}(x) - 1$, where $\hat{p}(x)$ is the probability to churn as estimated by the tree. When working with a proportional calibration sample, we set $\tau = 0$. In the presence of a nonproportional calibration sample, the value of τ varies (see next section).

From the original calibration set Z, we construct B bootstrap samples Z_b^*, $b = 1, 2, \ldots, B$ by randomly drawing, with replacement, N observations from Z. Note that the size of the bootstrap samples equals the original calibration sample size. From each bootstrap sample Z_b^*, we estimate a base classifier, giving B score functions $\hat{f}_1^*(x), \ldots, \hat{f}_b^*(x), \ldots, \hat{f}_B^*(x)$. We aggregate these functions into the final score

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}_b^*(x). \tag{1.2}$$

We can then carry out the classification using the following equation:

$$\hat{c}_{bag}(x) = sign(\hat{f}_{bag}(x) - \tau_B), \text{ with } \hat{c}_{bag}(x) \in \{-1, 1\}. \tag{1.3}$$

Again, the cutoff value τ_B equals zero in the presence of a proportional calibration sample. To determine the optimal value of B (i.e. the number of bootstrap samples), a strategy consists of selecting B such that the apparent error rates[4] (i.e. error rates on the calibration data) remain more or less constant for values larger than B. In our application, we set $B = 100$.

As with traditional classification models, we can also obtain diagnostic measures for the estimated bagging model. These are important to give

[4] Other criteria could also be considered (e.g. the Gini coefficient, the top-decile lift).

some face validity to the estimated model. For example, the estimated relative importance of each predictor in the construction of the classification rule can be investigated. For a single tree, the relative importance of a predictor can be computed, as Hastie, Tibshirani and Friedman (2001) do.[5] For bagging (and similarly for boosting), the relative importance of an explicative variable is averaged over all B trees. In addition, the partial dependence of churn on a specified predictor variable can be investigated. This measure provides similar insight to the parameter estimates' values of a logit model, but it advantageously allows for non-linear relationships between the predictors and the dependent variable. A partial dependence plot represents the impact of a predictor variable on the churn probability of a customer, conditional on all other predictors. In practice, the partial dependence of the dependent variable on a specified value of a predictor x_k is obtained by assigning this value of x_k to all observations of the calibration sample. The model is subsequently estimated,[6] and the N resulting predicted probabilities are computed for the calibration data. The partial dependence on a specified value of x_k is eventually given by averaging over these N predicted probabilities. The partial dependence *plot* is obtained by letting the value assigned to x_k vary over a large range of values (for more details, see Friedman 2001).

1.3.2 Boosting and Stochastic Gradient Boosting

Several versions of boosting exist, e.g. the Real AdaBoost (Freund and Schapire 1996; Schapire and Singer 1999), LogitBoost (Friedman, Hastie

[5] More precisely, a tree is composed of several nodes, from the root to the leaves (i.e. terminal nodes). Each nonterminal node is split into two child nodes on the basis of the value of the variable that provides the maximal reduction in the squared error rate. The relative importance of a variable x_k is then the sum of these improvements (reductions) over all nodes for which the predictor x_k was selected as a splitting variable.

[6] The estimation of the model on the *modified* values of x_k is only performed to construct a partial dependence plot, and the resulting probabilities are not used for predictive purpose.

and Tibshirani 2000), and gradient boosting (Friedman 2001). Boosting is more complex than bagging and not as easy to put into practice. In this chapter, we focus on stochastic gradient boosting (Friedman 2002), one of the most recent boosting variants and the winning model of the Teradata Churn Modeling Tournament (Cardell, Golovnya and Steinberg 2003).

The main difference between boosting and the previously described bagging procedure basically lies in the sampling scheme. Boosting consists of sequentially estimating a classifier to *adaptively reweighted* versions of the initial calibration sample Z_b^*, $b = 1, 2, \ldots, B$. The adaptive reweighting scheme enables us to give previously misclassified customers an increased weight on the next iteration, whereas weights given to observations that were correctly classified previously are reduced. The idea is to force the classification procedure to concentrate on the customers that are difficult to classify.

Another main difference with bagging is that the initial choice model should preferably be *weak* (i.e. with a slightly lower associated error rate than random guessing). For stochastic gradient boosting, Friedman (2002) recommends the use of k-node trees as a base classifier, where k is approximately 6-9, depending on the issue. In addition, the number of required iterations is usually higher for stochastic gradient boosting than for bagging. In our application, we select $B = 1000$.

1.4 Correction for a Balanced Sampling Scheme

Predictions made from a model estimated on a *balanced* calibration sample are known to be biased because they overestimate the proportion of churners in real life. Although appropriate bias correction methods already exist for some common classifiers (for the logit model, see e.g. King and Zeng 2001b), to the best of our knowledge, no correction method for bagging and boosting currently exists. Hereinafter, we adapt to the bagging and boosting models two simple bias correction methods that King and Zeng (2001b) discussed. The first correction consists of attaching a weight to each observation of the balanced sample. These weights are based on marketers' prior beliefs about the churn rate π_c (i.e. the proportion of churners) among their customers. For example, π_c can be taken as the empirical frequency of churners in a proportional sample; in the current context, this is 1.8%. Let $N_c^{balanced}$ be the number of churners in the balanced sample, where N is the total size of this sample. It is possible to weight the observations of a balanced calibration sample by attaching the weights

$$w_i^c = \frac{\pi_c}{N_c^{balanced}} \text{ and } w_i^{nc} = \frac{1-\pi_c}{N - N_c^{balanced}} \tag{1.4}$$

to the churners and the nonchurners, respectively. As such, the sum of the weights associated with the churners equals the real-life proportion of churners. Note that the sum of the weights we defined in Equation 1.4 is always equal to one. When this weighting correction is applied to bagging and stochastic gradient boosting, a sequence of weighted decision trees is estimated, and the weights remain fixed through iterations. In a statistical context, assigning weights to customers is a valid approach to correct for stratified sampling. However, because the weights assigned to the churners are small, this correction might actually cancel the advantage of oversampling the churners and thus provide similar results to a proportional sample of the same size (see the Results section).

Rather than weighting the observations of a balanced sample, we could employ a more simple approach by taking a nonzero cutoff value τ_B in the bagging and boosting algorithms. The value of τ_B is such that the proportion of predicted churners in the calibration sample equals the actual a priori proportion of churners π_c. This correction is achieved for bagging (and similarly for boosting) by first sorting the values of $\hat{f}_{bag}(x)$ in the calibration sample from the largest to the smallest value, $\hat{f}_{bag}(x_{(1)}) \geq \hat{f}_{bag}(x_{(2)}) \geq \ldots \geq \hat{f}_{bag}(x_{(N)})$, and then taking

$$\tau_B = \hat{f}_{bag}(x_{(j)}), \text{ where } j = N\pi_c. \tag{1.5}$$

This latter correction method can also be called *intercept correction* (or *prior correction* as in King and Zeng 2001a,b), referring to a similar correction for the logit model (see e.g. Franses and Paap 2001, pp. 73-75). Unlike the weighting correction, the intercept correction affects neither the estimated scores nor the ranking of the customers. We assess both corrections in the Results section.

1.5 Assessment Criteria

We assess the predictive performance of the investigated models using a holdout test sample (as described in the Data section). Because this sample has not been used for the estimation of the classification rules and is very large, it allows for a valid assessment of performance. We denote the validation or holdout test sample as $\{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_M, y_M)\}$, and the computed scores as $\hat{f}(x_i)$, for $i = 1, \ldots, M$, where M is the size of the validation sample.

1.5.1 Error Rate

The traditional performance criterion is the error rate, that is, the percentage of incorrectly classified observations in the validation set. For rare events, as Morrison (1969) notes, the error rate is often inappropriate. For example, a naive prediction rule stating that no customer of the validation set churns has an expected error rate of approximately 1.8%, from which the classification rule could be falsely considered good. Indeed, such a rule does not isolate any group of the potentially riskiest customers for a targeted retention strategy. Another drawback is that error rates do not take the numerical values of the scores $\hat{f}(x_i)$ into account, whereas these scores may contain relevant information for proactive marketing actions. The targeting of such incentives can indeed be based on the churn degree of risk (i.e. score) of each customer (e.g. targeting the 10% riskiest customers). In contrast, the top-decile lift and the Gini coefficient are based on these scores.

1.5.2 Top-Decile Lift

The top-decile lift focuses on the most critical group of customers and their churn risk. The top 10% riskiest customers (i.e. those who have score values among the 10% highest) represent a potentially ideal segment for targeting a retention marketing campaign. The top-decile lift equals the proportion of churners in this *risky* segment, $\pi_{10\%}$, divided by the proportion of churners in the whole validation set, π:

$$Top\ Decile = \frac{\hat{\pi}_{10\%}}{\hat{\pi}}. \tag{1.6}$$

The higher the top-decile lift, the better is the classifier. This measure enables us to control whether the targeted segment of risky customers indeed contains actual churners. As Neslin and colleagues (2004) extensively describe, top-decile lift is related directly to profitability. They define the incremental gain in financial profit from an increase in top-decile lift as

$$Gain = N\alpha\hat{\pi}(\Delta Top\ Decile)(\gamma LVC - \delta(\gamma - \psi)), \tag{1.7}$$

where N is the customer base of the company, α is the percentage of targeted customers (in our context, 10%), $\Delta Top\ Decile$ is the increase in top-decile lift, γ is the success rate of the incentive among the churners, LVC is the lifetime value of a customer (Gupta, Lehmann and Stuart 2004), δ is the incentive cost per customer, and ψ is the success rate of the incentive among the nonchurners (for more details, see Neslin et al. 2004).

1.5.3 Gini Coefficient

Another interesting measure is the Gini coefficient (e.g. Hand 1997, p.134). Instead of focusing only on the riskiest segment, this measure considers all scores, including the less risky customers. The top-decile lift and the Gini coefficient provide complementary information; a model can be good at identifying the riskiest segment but less effective at recognizing less risky customers. We first determine the fraction of *all subscribers* who have a predicted churn probability above a certain threshold. We consider a whole sequence of thresholds, each of which is given by a predicted score $\hat{f}(x_l)$, for $l = 1, 2, \ldots, M$, which results in M proportions

$$\pi_l = \frac{1}{M} \sum_{i=1}^{M} I[\hat{f}(x_i) > \hat{f}(x_l)]. \tag{1.8}$$

For each threshold, we also compute the fraction of *all churners* who have a score value above this threshold

$$\pi'_l = \frac{1}{M_c} \sum_{i=1}^{M} I[\hat{f}(x_i) > \hat{f}(x_l) \text{ and } y_i = 1], \tag{1.9}$$

where M_c is the total number of actual churners in the validation set. We then defined the Gini coefficient as

$$Gini\ \ Coefficient = \frac{2}{M} \sum_{l=1}^{M} (\pi'_l - \pi_l). \tag{1.10}$$

The larger the Gini coefficient, the better is the classification model.

1.6 Results

This section addresses the research questions introduced in the beginning of the chapter. We show that (i) both bagging and boosting techniques significantly improve the classification performance of traditional classification models, (ii) the correction methods for a balanced calibration sample reduce the classification error rate, and (iii) the use of a balanced calibration sample improves the forecasting accuracy of the estimated choice models.

1.6.1 Do Bagging and Boosting Provide Better Results than other Benchmarks?

We apply bagging and stochastic gradient boosting - with classification trees as base classifiers - to the balanced calibration sample.[7] As a benchmark, we estimate a binary logit choice model on the same sample. Other benchmark models, including the traditional discriminant analysis, a single classification tree, and a neural network, have also been investigated (see e.g. Thieme, Song and Calantone 2000; West, Brockett and Golden 1997), but they appear to perform worse than the binary logit choice model in this empirical application. Neslin and colleagues (2004) recently compared the predictive performance of different methodological approaches for this particular database and found that the logit model and the decision tree were among the most competitive methodologies. To evaluate the relative performance of the different methods, we apply the estimated models to the holdout proportional test sample to obtain churn predictions for each of the customers who belong to this sample. From these predictions, we then compute the validated error rate, the Gini coefficient and the top-decile lift that each of the three choice models reached.

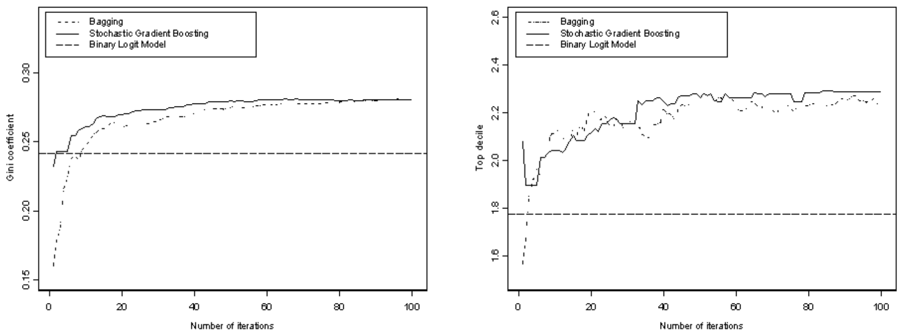
Figure 1.1 represents the Gini coefficient and the top-decile lift against the number of iterations for both bagging and stochastic gradient boosting without bias correction (which are identical to the results with intercept correction).[8] The horizontal line in Figure 1.1 represents the performance of the binary logit model. The performance of bagging and boosting improves as B increases and stabilizes for large values of B. After the first few iterations, both models already outperform the logit benchmark,[9] thus

[7]We implemented bagging using the statistical software package Splus, whereas we computed stochastic gradient boosting using the MART software package for R that J.H. Friedman developed.

[8]Note that B is actually multiplied by ten for stochastic gradient boosting in Figure 1.1.

[9]The Gini coefficient and top-decile lift are -.06 and .49, respectively, for neural nets;

Figure 1.1: *Validated Gini coefficient (left) and top-decile lift (right) for bagging, stochastic gradient boosting, and a binary logit model as a function of B.*



confirming many other examples (e.g. Hastie, Tibshirani and Friedman 2001, pp.246-249 & 299-345).

The relative gain in predictive performance is greater than 16% for the Gini coefficient, and 26% for the top-decile lift. This improvement is statistically significant.[10] Stochastic gradient boosting performs similarly to bagging but is conceptually more complicated. Therefore, we consider bagging the most competitive approach, at least in this application. We can also evaluate the additional financial gains (Equation 1.7) expected from a retention marketing campaign that would be targeted using the scores predicted by the bagging rather than the logit model. If we consider $N = 5,000,000$ customers, a target group of $\alpha = 10\%$, $\gamma = 30\%$ success

.199 and 1.60, respectively, for discriminant analysis; and 0.091 and 1.37, respectively, for a single classification tree, compared with .24 and 1.77 for logit regression. These figures motivate our preference for the logit model as a benchmark.

[10]Standard errors for both bagging and stochastic gradient boosting are approximately .012 for the Gini coefficient and .09 for the top-decile lift. Standard errors for the Gini coefficient and the top-decile lift are computed by drawing 1000 bootstrapped samples from the predicted probabilities. These samples yield 1000 replicates of the Gini coefficient or top-decile lift, from which the standard deviations are computed.

probability among the churners, $LVC = \$2,500$ lifetime value, $\delta = \$50$ incentive cost, and $\psi = 50\%$ success probability among the nonchurners, the use of bagging as a scoring model (versus a logit model) for targeting a specific retention campaign is worth an additional \$3,214,800. Using stochastic gradient boosting would yield another \$300,000 to be added the financial gains when using bagging.

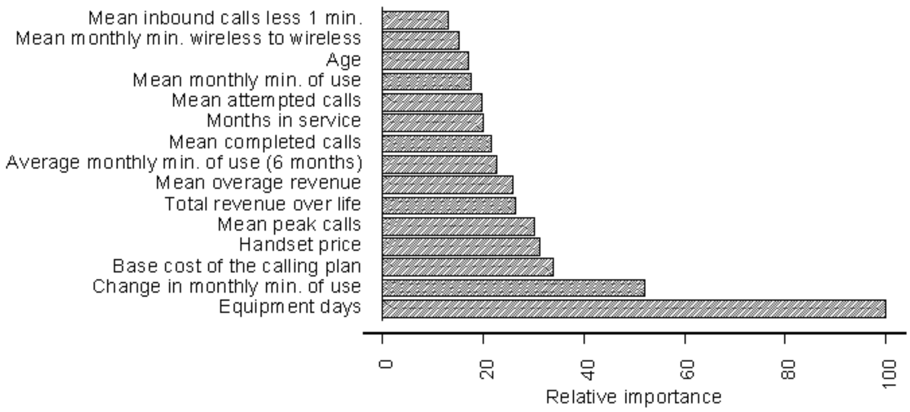
Regarding the error rate, all three choice models perform poorly (see Table 1.2; third column), confirming that a balanced sampling scheme requires an appropriate bias correction, regardless of the choice model under consideration. In the next research question, we investigate whether a bias correction reduces these high error rates.

Although the bagging and boosting models focus mainly on scoring customers for targeting purposes, we can also interpret the models. Figure 1.2 reports the 15 most important variables in explaining churn, using bagging.[11] Reported results offer some face validity. Among the particularly relevant churn triggers, we find the number of days of the current cellular phone ("equipment days"), the changes in minutes of consumption over the previous three months ("change in monthly minutes of use"), and the base cost of the calling plan the customer chose ("base cost of the calling plan"). Partial dependence plots provide additional insights into the way these variables affect churn.

It appears that (Figure 1.3; right panel) the probability that a customer churns increases as his or her cellular phone becomes older. This rise is particularly important during the first year, which could be due to numerous operators proposing combined one-year-subscription and free cellular phone packages. After this delay, customers may be likely to defect from the company and buy a new package from a competitor. Figure 1.3 (left panel) indicates how the churn risk of a customer varies as his or her consumption habits change. When consumption decreases, the subscriber

[11] Boosting yields similar results, confirming the face validity of the results.

Figure 1.2: *Variables' relative importance for bagging.*

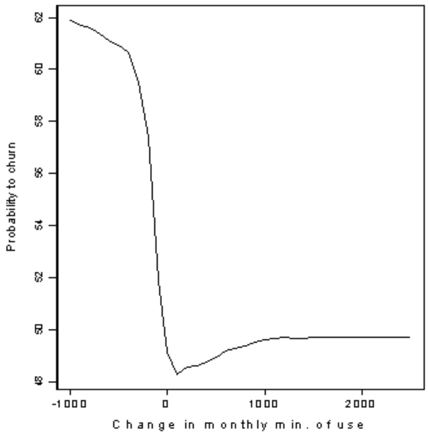


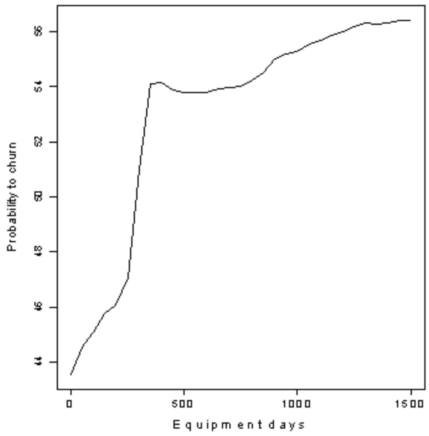
is more likely to churn. When his or her consumption is constant, the subscriber is less likely to defect. Finally, when consumption increases, the customer is slightly less (but still) likely to be loyal than when no change occurs.[12]

Another interesting insight can be derived from Figure 1.4, which represents the partial dependence between churn and a combination of two churn drivers, i.e. the age of the customer ("age") and the base cost of his or her calling plan. A customer is found to be more likely to churn when his or her calling plan is cheaper. However, this relationship tends to be much stronger for younger customers than for older ones, indicating that some demographics are more likely to drop certain calling plans than others.

[12]Note that logit models cannot capture such nonmonotonic relations.

Figure 1.3: *Partial dependence plots for "change in monthly minutes of use" and "equipment days" for bagging.*





1.6.2 What is the Best Bias Correction when Using a Balanced Calibration Sample?

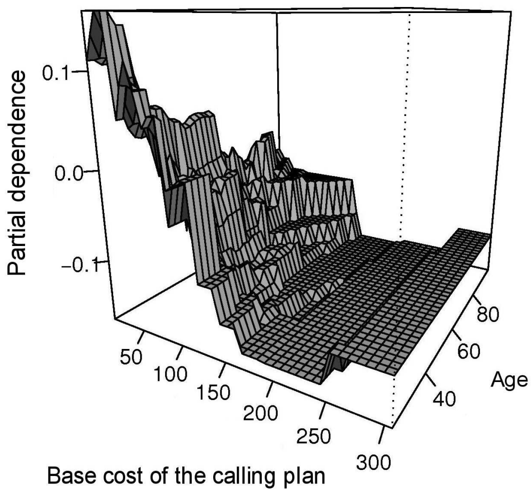
We use two corrections to adapt the predicted probabilities obtained through the use of a balanced calibration sample. Either of these two corrections reduces the error rate significantly, as illustrated in Table 1.2.

Table 1.2: *Validated error for predicting churn from a balanced sample with intercept correction, with weighting correction, or without bias correction.*

Error Rate	Intercept Correction	Weighting Correction	No Correction
Binary logit model	0.035	0.018	0.400
Bagging	0.034	0.025	0.374
Stochastic gradient boosting	0.034	0.018	0.460

The effectiveness of both corrections differs. For the error rate, the weighting correction seems to be the most appropriate bias correction

Figure 1.4: *Partial dependence plot for the "base cost of the calling plan" and "age" for bagging.*



method for all considered models. However, the weighting correction affects the estimated scores, their ranking, and, eventually, the Gini coefficient and the top-decile lift. This is not the case for the intercept correction method which preserves the relative ranking of the attributed scores. Table 1.3 reports the Gini coefficient and the top-decile lift for bagging, stochastic gradient boosting, and the logit model (all estimated on the balanced sample) for both corrections. For all the three models under consideration, the Gini coefficient and the top-decile lift obtained with the intercept correction are substantially better than those obtained with the weighting correction. The weighting correction turns out to be particularly inefficient for bagging. This low performance can be related to the fact that the weights are included in the bagging iterations, which therefore removes part of the randomization in the bootstrap sampling. These results confirm the prior assumption that weighting the observations of a balanced sample cancels the advantage of balanced sampling, even for large sample sizes. Because we consider the Gini coefficient and the top-decile lift more global measures of performance than the error rate, the intercept correction is found to be the best compromise between no correction (i.e. a better Gini coefficient and top-decile lift, but a worse error rate) and weighting correction (i.e. a worse Gini coefficient and top-decile lift, but a better error rate), at least in this application.

Note that the intercept correction appears to perform well for stable markets (e.g. constant churn rate), but it is likely to be inefficient in dynamic markets (e.g. increasing churn rate). This constitutes a major limitation to the correction methods we propose in this study. Moreover, the lack of theory about the properties of these correction methods prevents us from generalizing our findings to any other setting.

Table 1.3: *Validated Gini coefficient and top-decile lift for predicting churn from a balanced sample with intercept correction and weighting correction.*

	Intercept Correction[a]		Weighting Correction	
	Gini Coefficient	Top Decile	Gini Coefficient	Top Decile
Binary logit model	0.241	1.775	0.239	1.764
Bagging	0.281	2.246	0.161	1.549
Stochastic gradient boosting	0.280	2.290	0.187	1.632

[a] The Gini coefficients and top-decile lifts are the same for the "no correction" method.

1.6.3 Does a Choice Model Estimated on a Balanced Sample, with Bias Appropriately Corrected for, Outperform a Choice Model Estimated on a Proportional Sample?

A balanced calibration sample is often advised when the variable to be predicted consists of a rare event, such as churn. However, our third research issue questions this advice. Indeed, given the high amount of observations in the proportional calibration sample, the absolute number of churners is still quite large, and a proportional sampling could still be efficient.

Table 1.4 compares the performance of bagging, stochastic gradient boosting, and the binary logit model, estimated from the proportional or the balanced sample (with intercept correction). The results of both the Gini coefficient and the top-decile lift indicate that the balanced sampling scheme is recommended for the three investigated classification models. For the error rate, the results are more in favor of proportional sampling. However, for the same reasons as in the preceding subsection, we consider the balanced sampling a better compromise than the proportional sampling, which performs poorly for the Gini coefficient and top-decile lift.

1.7 Conclusions

In this chapter, we discussed several new developments from the machine-learning and statistical classification literature in the context of marketing research. We presented one of the simplest versions of classifier aggregation, i.e. bagging, and one of the most sophisticated algorithms in this field, i.e. stochastic gradient boosting. We especially drew attention to the competitive performance of bagging, an easy-to-use procedure aimed at increasing the classification performance of an initial classification model, by repeatedly estimating a classifier to bootstrapped versions of the calibra-

Table 1.4: *Validated Gini coefficient, top-decile lift, and error rate with a balanced and a proportional calibration sampling.*

	Balanced Sample (Intercept Correction)			Proportional Sample		
	Gini Coefficient	Top Decile	Error Rate	Gini Coefficient	Top Decile	Error Rate
Binary logit model	0.241	1.775	0.035	0.181	1.665	0.018
Bagging	0.281	2.246	0.034	0.237	1.886	0.018
Stochastic gradient boosting	0.280	2.290	0.034	0.113	1.560	0.018

tion sample. We summarize the main findings of this study in terms of three contributions. First, bagging and boosting provide substantially better classifiers than a binary logit model. In predicting churn, the gain in predictive performance has reached 16% for the Gini coefficient and 26% for the top-decile lift. Bagging and stochastic gradient boosting perform comparatively. The performance of the simple and easy-to-use bagging is especially noticeable. In addition to their higher predictive power, bagging and boosting provide good diagnostic measures, variable importance, and partial dependence plots, which offer face validity to the models and interesting insights into potential churn drivers.

Second, in the presence of a rare event, such as churn, we recommend a balanced sampling scheme over proportional sampling for all considered classification models (i.e. bagging, boosting and logit models), even for large data sets. However, to maintain the classification error rate at a reasonable level, it is necessary to correct the predictions obtained from a balanced sample. Third, intercept correction constitutes an appropriate bias correction for a balanced sampling scheme.

If companies take into account these recommendations, they might better identify the riskiest customer segments in terms of churn risk and thus ameliorate their retention strategy. Noteworthy losses could ultimately be avoided.

References

[1] Andrews, Rick L., Andrew Ainslie, and Imran S. Currim (2002), "An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity," *Journal of Marketing Research*, 39 (November), 479-87.

[2] — and Imran S. Currim (2002), "Identifying Segments with Identical Choice Behaviors Across Product Categories: An Intercategory Logit Mixture Model," *International Journal of Research in Marketing*, 19 (1), 65-79.

[3] Arora, Neeraj, Greg M. Allenby, and James L. Ginter (1998), "A Hierarchical Bayes Model of Primary and Secondary Demand," *Marketing Science*, 17 (1), 29-44.

[4] Athanassopoulos, Antreas D. (2000), "Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behavior," *Journal of Business Research*, 47 (3), 191-207.

[5] Baines, Paul R., Robert M. Worcester, Jarrett David, and Roger Mortimore (2003), "Market Segmentation and Product Differentiation in Political Campaigns: A Technical Feature Perspective," *Journal of Marketing Management*, 19 (1-2), 225-49.

[6] Bhattacharya, C.B. (1998), "When Customers are Members: Customer Retention in Paid Membership Contexts," *Journal of the Academy of Marketing Science*, 26 (1), 31-44.

[7] Bolton, Ruth N., P.K. Kannan, and Matthew D. Bramlett (2000), "Implications of Loyalty Program Membership and Service Experiences for Customer Retention and Value," *Journal of the Academy of Marketing Science*, 28 (1), 95-108.

[8] Breiman, Leo (1996), "Bagging Predictors," *Machine Learning*, 24 (2), 123-40.

[9] —, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone (1984), *Classification and Regression Trees*. New York: Chapman and Hall.

[10] Bühlmann, Peter and Bin Yu (2002), "Analyzing Bagging," *Annals of Statistics*, 30 (4), 927-61.

[11] Cardell, Scott N., Mikhail Golovnya, and Dan Steinberg (2003), "Churn Modeling for Mobile Telecommunications: Winning the NCR Teradata Center for CRM at Duke University: Salford Systems," paper presented at the 2003 INFORMS Marketing Science Conference, University of Maryland (June 12-15).

[12] Chung, Jaihak and Vithala R. Rao (2003), "A General Choice Model for Bundles with Multiple-Category Products: Application to Market Segmentation and Optimal Pricing for Bundles," *Journal of Marketing Research*, 40 (May), 115-30.

[13] Colgate, Mark R. and Peter J. Danaher (2000), "Implementing a Customer Relationship Strategy: The Asymmetric Impact of Poor Versus Excellent Execution," *Journal of the Academy of Marketing Science*, 28 (3), 375-87.

[14] Corstjens, Marcel L. and David A. Gautschi (1983), "Formal Choice Models in Marketing," *Marketing Science*, 2 (1), 19-56.

[15] Cosslett, S.R. (1993), "Estimation from Endogenously Stratified Samples," in *Handbook of Statistics*, Vol. 11, G.S. Maddala, C.R. Rao, and H.D. Vinod, eds. Amsterdam: Elsevier Science.

[16] Currim, Imran S., Robert J. Meyer, and Nhan T. Le (1988), "Disaggregate Tree-Structure Modeling of Consumer Choice Data," *Journal of Marketing Research*, 25 (August), 253-65.

[17] Donkers, Bas, Philip H.B.F. Franses, and Peter Verhoef (2003), "Selective Sampling for Binary Choice Models," *Journal of Marketing Research*, 40 (November), 492-97.

[18] —, Richard Paap, Jedid-Jah Jonker, and Philip H.B.F. Franses (2006), "Deriving Target Selection Rules from Endogenously Selected Samples," *Journal of Applied Econometrics*, forthcoming.

[19] Franses, Philip H. and Richard Paap (2001), *Quantitative Models for Marketing Research.* Cambridge, UK: Cambridge University Press.

[20] Freund, Yoav and Robert E. Schapire (1996), “Experiments with a New Boosting Algorithm,” in *Proceedings of the 13th International Conference on Machine Learning*, L. Saitta, ed. Bari, Italy: Morgan Kaufmann, 148-56.

[21] Friedman, Jerome H. (2001), “Greedy Function Approximation: A Gradient Boosting Machine,” *Annals of Statistics*, 29 (5), 1189-1232.

[22] — (2002), “Stochastic Gradient Boosting,” *Computational Statistics and Data Analysis*, 38 (4), 367-78.

[23] —, Trevor Hastie, and Robert Tibshirani (2000), “Additive Logistic Regression: A Statistical View of Boosting,” *Annals of Statistics*, 28 (2), 337-407.

[24] Ganesh, Jaishankar, Mark J. Arnold, and Kristy E. Reynolds (2000), “Understanding the Customer Base of Service Providers: An Examination of the Differences Between Switchers and Stayers,” *Journal of Marketing*, 64 (July), 65-87.

[25] Guadagni, Peter M. and John D.C. Little (1983), “A Logit Model of Brand Choice Calibrated on Scanner Data,” *Marketing Science*, 2 (3), 203-238.

[26] Gupta, Sunil, Donald R. Lehmann, and Jennifer A. Stuart (2004), “Valuing Customers,” *Journal of Marketing Research*, 41 (February), 7-18.

[27] Hand, David J. (1997), *Construction and Assessment of Classification Rules.* Chichester, UK: John Wiley & Sons.

[28] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2001), *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* New York: Springer-Verlag.

[29] Haughton, Dominique and Samer Oulabi (1997), “Direct Marketing Modeling with CART and CHAID,” *Journal of Direct Marketing*, 11 (4), 42-52.

[30] Hawley, David (2003), “International Wireless Churn Management: Research and Recommendations,” Yankee Group report, (June), (accessed January 2006), [available at http://www.ams.com/cme/pdfs/yankeechurnstudy.pdf].

[31] Imbens, Guido W. and Tony Lancaster (1996), “Efficient Estimation and Stratified Sampling,” *Journal of Econometrics*, 74 (2), 289-318.

[32] Kalwani, Manohar U., Robert J. Meyer, and Donald G. Morrison (1994), "Benchmarks for Discrete Choice Models," *Journal of Marketing Research*, 31 (February), 65-75.

[33] King, Gary and Langsche Zeng (2001a), "Explaining Rare Events in International Relations," *International Organization*, 55 (3), 693-715.

[34] — and — (2001b), "Logistic Regression in Rare Events Data," *Political Analysis*, 9 (2), 137-63.

[35] Morrison, Donald G. (1969), "On the Interpretability of Discriminant Analysis," *Journal of Marketing Research*, 6 (May), 156-63.

[36] Nardiello, Pio, Fabrizio Sebastiani, and Alessandro Sperduti (2003), "Discretizing Continuous Attributes in AdaBoost for Text Categorization," in *Proceedings of the 25th European Conference on Information Retrieval Research*, Fabrizio Sebastiani, ed. Heidelberg, Germany: Springer-Verlag, 320-34.

[37] Neslin, Scott A., Sunil Gupta, Wagner Kamakura, Junxiang Lu, and Charlotte Mason (2004), "Defection Detection: Improving Predictive Accuracy of Customer Churn Models," working paper, Teradata Center for Customer Relationship Management, Duke University.

[38] Schapire, Robert E. and Yoram Singer (1999), "Improved Boosting Algorithms Using Confidence-Rated Predictions," *Machine Learning*, 37 (3), 297-336.

[39] Scott, Alastair J. and Chris J. Wild (1997), "Fitting Regression Models to Case-Control Data by Maximum Likelihood," *Biometrika*, 84 (1), 57-71.

[40] Shaffer, Greg and John Z. Zhang (2002), "Competitive One-to-One Promotions," *Management Science*, 48 (9), 1143-60.

[41] Snel, Ross (2000), "Fighting the Fickle," *The Wall Street Journal (Europe)*, (September 18), (accessed January 2006), [available at http://interactive.wsj.com/public/current/articles /SB968956754675902638.htm].

[42] Thieme, R. Jeffrey, Michael Song, and Roger J. Calantone (2000), "Artificial Neural Network Decision Support Systems for New Product Development Project Selection," *Journal of Marketing Research*, 37 (November), 499-507.

[43] Varmuza, Kurt, Ping He, and Kai-Tai Fang (2003), "Boosting Applied to Classification of Mass Spectral Data," *Journal of Data Science*, 1 (4), 391-404.

[44] Viaene, Stijn, Richard A. Derrig, and Guido Dedene (2002), "Boosting Naive Bayes for Claim Fraud Diagnosis," in *Lecture Notes in Computer Science 2454*, Y. Kambayashi, W. Winiwarter, and M. Arikawa, eds. Berlin: Springer, 202-211.

[45] Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, 2d ed. Boston: Kluwer Academic Publishers.

[46] West, Patricia M., Patrick L. Brockett, and Linda L. Golden (1997), "A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice," *Marketing Science*, 16 (4), 370-91.

[47] Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40 (August), 282-94.

Chapter 2

Bagging a Stacked Classifier

2.1 Introduction

Suppose that one needs to classify an email as being spam or not. In this perspective, several supervised two-class classification methods (e.g. logistic regression, k-nearest neighbors, neural networks, decision trees, discriminant analysis, etc.) can be used. However, one could quickly get lost and waste time to find out the best classifier. Each classifier has indeed its own advantages and disadvantages and could well be adapted to one specific task, but not recommended for another. This issue can be solved by combining, in a fully-automated way, different classifiers. The combined classifier should perform better than each single classifier.

In the classification literature, several classifier combination methods have already been proposed and can be categorized in two groups. Some algorithms combine classifiers of the same class. For example, the boosted tree is a linear combination of several decision trees (e.g. Hastie et al 2001). A second category of combination methods handles heterogeneous classifiers. Within this class, we focus on a method called stacked generalization, introduced by Wolpert (1992) in the neural network literature; also called stacked classification, by LeBlanc and Tibshirani (1996), in the statistical literature. The next section briefly reviews the stacking principle.

In this chapter, we propose a new, easy-to-implement and flexible algorithm to combine different classifiers such that the final error rate, as estimated by ten-fold cross-validation, is minimized by construction. Furthermore, in the third section of this chapter, we apply bagging to reduce the variability of the stacked classifier, and to further improve its performance.

2.2 Stacked Classification

Consider a calibration data set $Z = \{(x_i, y_i)|1 \leq i \leq n\}$, containing n observations. For the calibration data, we observe the values of the binary variable Y, indicating which source population the observation belongs to. This variable Y needs to be predicted using X, a vector of p explicative variables, which may be continuous or categorical. The realizations of X on the calibration set are $x_1, \ldots, x_n$. The aim is to predict the value of Y for a new instance x. A classifier $C(x; Z)$ is constructed on the basis of the calibration set. It takes values on the interval $[0, 1]$: the higher the value of $C(x; Z)$, the higher the likelihood that an observation with features x belongs to the source population with label "Y=1". The value of $C(x; Z)$ is called the "score" attached to x. Classifiers taking values 0 and 1 only can be seen as a special case were only the most extreme values are given as score.

Several supervised classification rules exist. In this chapter, four well-known methods are considered: classification trees, discriminant analysis, logistic regression and k-nearest neighbors. Classification trees yield a non-parametric classification procedure, and are widely used in applied fields such as medicine and botany. Linear discriminant analysis is the most traditional supervised classification method, optimal if both source populations are normal with the same covariance matrix. It is a parametric method, where the scores are given by linear combinations of the explicative variables. Logistic regression is another parametric classification technique where the conditional probabilities $P(Y = 1|x)$ are modeled by means of a logit transformation. Finally, the k-nearest neighbors method consists of finding the k-closest neighbors to an observation x (according to a certain distance). The value $C(x; Z)$ is given by the frequency of observations for which $y_i = 1$, among the k-nearest neighbors (we select $k = \sqrt{n}$). We refer to the standard S-plus implementations to compare the different methods under consideration. Of course, other classification methods could also have been used as level-zero classifiers.

It is well recognized that it does not exist a unique best classifier. Suppose that K different initial classifiers $C_1, \ldots, C_K$ are available. These initial classifiers can be combined in one single stacked classifier. Wolpert (1992) called the initial classifiers the "level-zero" components. The idea is to use the outcomes of the "level-zero" classifiers as inputs to find a "level-one" classifier. In other words, the calibration data set for this "level-one" classifier is given by $\{(C^*(x_i; Z), y_i)|1 \leq i \leq n\}$, where $C^*(x_i; Z) = (C_1(x_i; Z), \ldots, C_K(x_i; Z))^t$ is obtained by stacking the different level-zero classifiers. As level-one classifiers, one could again consider decision tree, logistic regression, etc.

Wolpert (1992) and Leblanc and Tibshirani (1996) advise to use predicted scores obtained by cross-validation instead of the $C_j(x_i, X)$, for $j = 1, \ldots, K$ and $i = 1, \ldots, n$. As such, less overfitting should occur, leading to a better performance on future data to classify. Moreover, the cross-validation reduces the risk of having potentially high correlation between the level-zero classifiers. Hence, throughout this chapter, the level-one classifiers will be constructed from the stacked vector $C^*(x_i; Z) = (C_1^-(x_i; Z), \ldots, C_K^-(x_i; Z))^t$, where $C_j^-(x_i; X)$ is obtained via cross-validation for $j = 1, \ldots, K$. To reduce the computation time, the $C_j^-(x_i; Z)$ were computed via 10-fold cross-validation. Formally, the calibration data set Z is split into ten subsets, Z_1, ..., Z_{10}. Predictions for the observations belonging to a subset Z_l are then based on a classifier constructed with $Z \setminus Z_l$ as calibration set, for $l = 1, \ldots, 10$. The algorithm of an "optimal" level-one classifier is introduced in the next subsection.

2.2.1 A New Algorithm for Combining Classifiers

The idea behind stacking is to find an optimal linear combination of the component classifiers such that the following stacked classifier is obtained

$$C_W(x; Z) = \beta_1 C_1(x; Z) + ... + \beta_K C_K(x; Z). \tag{2.1}$$

The coefficients $\beta_1, \ldots, \beta_K$ need to be chosen in such a way that the stacked classifier $C_W(x; Z)$ performs "better" than the initial classifiers. We require these coefficients to sum up to one and to be non-negative, such that they can be interpreted as the weights attached to each classifier. If a component classifier performs particularly bad, it receives a low or even zero weight, and has no impact on the final classification. The combined classifier C_W is therefore an "optimal" *weighted* average of the component classifiers. The weights $\beta_1, \ldots, \beta_K$ are obtained in order to optimize a certain criterion, commonly the *error rate*, i.e. the percentage of misclassified observations. Other criteria do exist, such as the area under the receiver operating curve (ROC). The algorithm proposed in this chapter works for any criterion of choice, and can be used for any set of component classifiers, making it very flexible. In this short chapter, we restrict the attention on the error rate as performance criterion. Note that Leblanc and Tibshirani (1996) worked with the mean squared error as criterion, a natural choice in the regression setting but inappropriate in the classification literature. Since the population error rate is unknown, we work with the (10-fold) cross-validated error rate. This error rate is known to be unbiased, in contrast to the apparent error rate. Finding the optimal values for β is not easy, certainly not for high values of K. Hence, we propose the following greedy algorithm:

- Compute, via 10-fold cross validation, the scores for every component classifier $C_j^-(x_1; Z), \ldots, C_j^-(x_n; Z)$ for $j = 1, \ldots, K$.

- Using the previously computed scores, compute the cross-validated error rate associated with the classifiers (without prior information, classify an observation in the first source population if the value of the score exceeds the threshold 0.5). These are used to sort the classifiers from the smallest to the highest error rate. We denote the ordered classifiers $C_{(1)}, ..., C_{(K)}$.

- Maximizing the criterion over the space of all possible values of β_1, ..., β_K is difficult, since the objective function to minimize is not differentiable in general. We therefore propose an iterative procedure. First find α_1 such that the error rate of the combined classifier

$$C^1(x; Z) = \alpha_1 C_{(1)}(x; Z) + (1 - \alpha_1)\, C_{(2)}(x; Z)$$

is minimized, where α_1 ranges over the interval $[0, 1]$. A simple univariate optimization routine, like grid search, can be used. For $k = 2$ to $K - 1$, find α_k such that the error rate of

$$C^k(x; Z) = \alpha_k C^{k-1}(x; Z) + (1 - \alpha_k)\, C_{(k+1)}(x; Z)$$

is minimized. After a first cycle is completed, one could rearrange the terms and express C^{K-1} as a convex combination of all level-zero classifiers

$$\beta_1 C_1(x; Z) + ... + \beta_K C_K(x; Z).$$

A new cycle might then be computed, to successively find the best linear combination of $C_{(j)}$ and C^b, for $j = 1, \ldots, K$, where C^b is the currently "best" combination of level-zero classifiers. In the present version of the program, we perform three full cycles after which no significant further improvement was observed.

The coefficient $\beta_1, \ldots, \beta_K$ obtained in this way are used as weights for the final classifier C_W in (2.1). By construction, the classifier C_W always has a lower (10-fold cross-validated) error rate than each single level-zero classifier. Note that there is no guarantee, however, that the global optimum is reached.

Table 2.1: *Cross-validated error rates (in %) of the level-zero classifiers and the optimal weighted average C_W. The first two columns contain the number of observations n and number of variables p of each data set.*

			Level-Zero Classifiers				Stacked
	n	p	Tree	Disc	Logit	NN	C_W
Austral	690	14	17.03	13.59	12.50	28.80	11.78
Balloon	156	5	10.48	10.48	8.87	8.87	7.26
Breast	699	9	5.90	3.94	4.11	3.22	3.22
Cmc	1473	9	30.31	32.60	32.34	28.27	27.16
Crx	653	15	18.20	12.84	13.60	31.23	12.26
Iono	351	33	12.50	14.29	15.36	16.07	8.57
Mushroom	8124	21	0.15	6.03	3.46	4.39	0.06
Pimadiab	768	8	26.87	22.64	22.80	27.20	22.31
Spambase	4601	57	8.78	11.68	7.28	25.38	6.03
Tictacto	958	9	17.10	32.64	32.90	14.88	12.79
Wdbc	569	30	6.81	4.40	5.93	6.81	3.30
Wpbc	198	31	26.58	20.25	22.78	22.78	17.09

To illustrate this procedure, we applied the greedy algorithm for C_W to 12 data sets available on the UCI repository (Mertz and Murphy 1996). The size n and the number of explicative variables p of these data sets are reported in Table 2.1. The data sets are previously cleaned from missing values. Each data set is split into a calibration set Z (80% of the observations) and a test set T (20% of the observations). The calibration set Z is used to obtain both the level-zero classifiers and the stacked classifiers.

The test sets are used in the sequel.

Table 2.1 reports the cross-validated error rates for every level-zero classifier, i.e. classification tree ("Tree"), discriminant analysis ("Disc"), logistic regression ("Logit"), k-nearest neighbors ("NN") and for the optimal weighted average C_W of these four classifiers. No level-zero classifier clearly outperforms the others, justifying the choice for combining classifiers. As aforementioned, the error rates obtained by the C_W on the calibration sample are always lower than those associated with the initial classifiers. The improvement is quite substantial, amounting until 31.44 % (Iono data) or even 60.00 % (Mushroom data) of relative improvement with respect to the best level-zero classifier. In one case (the Breast data), the k-nearest neighbors classifier performs equally as C_W. In this case, the weight associated with the k-nearest neighbors classifier turns out to be one.

At first sight, the results are very encouraging. However, one should not forget that the weights of the combined classifier are determined to minimize the cross-validated error rate computed on the calibration data. There is a risk that the reported error rates, albeit obtained by cross-validation, are still over-optimistic. Therefore, in the following subsection, we compute the error rates on the *test set*, and compare the different level-one classifiers.

2.2.2 Empirical Comparison of Different Stacking Methods

Table 2.2 reports the error rates computed on the test set for the level-zero classifiers and the weighted average C_W. One may observe from Table 2.2 that the weighted method does not systematically provide the lowest test error rates, even if it is still the case for 8 out of 12 data sets. When a single component performs better, the weighted combination is closely comparable to the best solution. Moreover, for each level-zero classifier, it is possible to find at least one data set on which the level-zero classifier performs drastically worse than the stacked classifier C_W.

Table 2.2: *Test error rates in % for the 4 different level-zero classifiers, for the weighted classifier C_W, and for stacking by 4 different level-1 classifiers.*

	Level-Zero Classifiers				Stacked Classifiers				
	Tree	Disc	Logit	NN	C_W	Tree	Disc	Logit	NN
Austral	14.40	15.90	14.40	31.10	15.90	20.20	14.40	13.70	12.30
Balloon	9.38	9.38	9.38	9.38	9.38	12.50	12.50	9.38	9.38
Breast	6.43	9.29	8.57	6.43	6.43	4.29	6.43	5.71	7.14
Cmc	29.80	32.80	32.20	28.10	28.80	32.80	29.40	28.40	28.40
Crx	17.50	16.00	14.50	33.50	16.00	16.70	16.00	15.20	15.20
Iono	19.70	14.00	12.60	16.90	12.60	9.86	11.20	9.86	11.20
Mushroom	0.25	5.05	2.71	4.62	0.12	0.25	0.25	0.12	0.25
Pimadiab	25.90	20.70	20.70	21.40	20.70	27.20	21.40	22.00	20.70
Spambase	6.19	10.10	7.17	24.65	5.54	4.89	5.32	5.32	5.97
Tictacto	13.50	26.50	26.00	13.00	11.90	8.33	7.29	7.81	5.73
Wdbc	2.63	2.63	6.14	6.14	3.51	1.75	2.63	0.88	3.51
Wpbc	27.50	20.00	17.50	27.50	15.00	22.50	12.50	12.50	15.00

Table 2.2 also reports the results of other stacking methods. The level-one combination methods under consideration include a decision tree, linear discriminant analysis, logistic regression and k-nearest neighbors. One may observe that the different ways of combining the level-zero estimators yield different error rates, even if the heterogeneity is less pronounced than among the level-zero classifiers. The stacked logistic regression seems to be the best stacking method on the data sets under investigation, giving the smallest, or least one of the smallest, test error rates. Note that the weighted average does not outperform the other stacking methods on the test set. In the next section, we apply bagging on the stacked classifiers.

2.3 Bagging the Stacked Classifiers

Error rates estimated by cross-validation are known to be very volatile. The bagging technique, originating from machine learning, helps to reduce the variance and improves the performance of unstable classifiers (Breiman 1996). Therefore, we propose to apply the bagging algorithm to the weighted classifier C_W, described in the previous section. Practically, the bagging procedure consists of computing classifiers on B different bootstrap samples of size n drawn from the initial calibration set Z. These bootstrap samples are obtained by drawing with replacement observations from Z, yielding $Z^{*1}, \ldots, Z^{*b}$. The bagged classifier is then obtained by averaging over all predicted scores:

$$C_{bag,B}(x; Z) = \frac{1}{B} \sum_{b=1}^{B} C_W(x; Z^{*b}). \tag{2.2}$$

Bagging is an easy-to-implement procedure and can be applied to any given classification method. In Figure 2.1, the test error of $C_{bag,B}$ is plotted with respect to B, for $B = 1, \ldots, 50$. In the left panel, bagging is applied on the weighted average classifier C_W for the Cmc data set. One may observe that the test error of the bagged classifier improves the starting stacked classifier for large values of B. In the right panel of Figure 2.1, the performance gain obtained by bagging is illustrated even more clearly: the stacked classifier using discriminant analysis in level-one has a test error rate of 12.6%, decreasing until 8.14% after $B = 50$ bagging steps.

Figure 2.1: *Test error rates w.r.t. the number of bootstraps B for stacked classification using (left panel) the weighted average C_W for the Cmc data set (right panel) discriminant analysis for the Iono data set. The dotted line corresponds to the error rate for the classification method without bagging.*

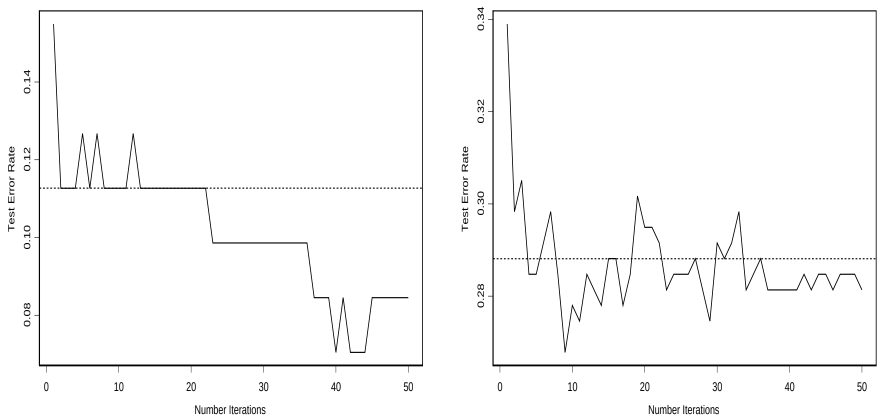


Table 2.3 reports the test error rates of the combination methods after 50 bagging steps. The results indicate that, for most data sets and classifiers, bagging indeed reduces the test error rate of the stacked classifiers (45 times out of 60). The bagging also reduces the differences in performance across the different stacking methods. Moreover, it turns out that the bagged weighted combination method outperforms the others on 6 out of 12 data sets, and is comparable to the bagged stacked classifiers using decision tree or logistic regression as level-one classifier.

2.4 Conclusions and Areas for Future Research

As a conclusion, this chapter introduced a new weighted combination method which was compared to existing stacking methods. It was shown that the weighted combination generally improves the performance of the

Table 2.3: *Test error rates of several bagged stacked classifiers (with $B = 50$): weighted average C_W together with stacking by 4 different level-1 classifiers.*

	Bagged Stacked Classifiers				
	C_W	Tree	Disc	Logit	NN
Austral	11.50	12.30	12.30	12.32	12.30
Balloon	9.38	9.38	12.50	9.38	12.50
Breast	5.71	4.29	5.00	5.00	6.43
Cmc	28.10	28.80	28.40	28.40	28.80
Crx	14.50	14.50	14.50	14.50	14.50
Iono	7.04	8.45	8.45	7.04	7.04
Mushroom	0.00	0.12	0.12	0.00	0.12
Pimadiab	21.40	22.00	20.70	20.10	23.30
Spambase	5.10	4.89	5.32	5.21	5.10
Tictacto	7.81	3.13	3.13	4.69	5.21
Wdbc	1.75	0.88	0.88	0.88	0.88
Wpbc	12.50	15.00	15.0	15.00	15.00

level-zero classifiers. Moreover, it performs well in comparison to the other stacking methods. Given that the variance of the stacked classifiers turned out to be high, we also investigated the opportunity to apply bagging on stacked classifiers. It turned out that bagging improves their performance. This performance gain is related to the fact that bagging reduces the high variance of the stacked classifiers. We therefore recommend to apply bagging after any stacking method of choice.

This piece of research opens several directions for future research. First of all, it would be insightful to compare the performance of the stacking algorithm proposed in this chapter with other existing methods. In particular, several classifier combination variants have been recently proposed in the field of (linear) mathematical programming (see e.g. Adem and

Gochet, 2006).

Another interesting area for future research concerns the use of the optimal weighting average algorithm for other performance criteria than the error rate. In particular in the churn context (see Chapter 1), it would be of high interest for marketers to use the top-decile lift as alternative performance criterion. As aforementioned, the latter has a direct interpretation in terms of financial gains that a targeted retention campaign can leverage (Neslin et al. 2004). As such, the use of a top-decile lift optimization method might ameliorate the design of marketing retention strategies.

References

[1] Adem, J. and Gochet, W. (2006), "Mathematical Programming Based Heuristics for Improving LP-Generated Classifiers for the Multiclass Supervised Classification Problem," *European Journal of Operational Research*, **168**, 181-199.

[2] Breiman, L. (1996), "Bagging Predictors," *Machine Learning*, **26**, 123-140.

[3] Hastie, T., Tibshirani, R. and Friedman, J. (2001), *The elements of statistical learning: data mining, inference, and prediction*, Springer-Verlag, New York.

[4] LeBlanc, M. and Tibshirani, R. (1996), "Combining estimates in regression and classification," *Journal of the American Statistical Association*, **91**, 1641-1647.

[5] Mertz, C.J. and Murphy, P.H. (1996), University of California, Irvine, Repository of machine learning databases (accessed January 2006), [available at http://www.ics.uci.edu/ mlearn/MLRepository.html].

[6] Neslin, S.A., Gupta, S., Kamakura, W., Lu, J., and Mason C. (2004), "Defection Detection: Improving Predictive Accuracy of Customer Churn Models," working paper, Teradata Center for Customer Relationship Management, Duke University.

[7] Wolpert, D.H. (1992), Stacked generalization, *Neural Networks*, **5**, 241-259.

Part II

Advanced Time-Series and Spectral-Based Models and Marketing Issues

Chapter 3

On the Predictive Content of Production Surveys: A Pan-European Study

3.1 Introduction

For over forty years, Business Tendency Surveys have been collected by the European Union in the framework of the Joint Harmonised EU Programme of Business and Consumer Surveys. The Programme was initially set up in 1961 by the European Commission, and currently includes the fifteen member states of the European Union. Business Surveys are carried out on a monthly basis by various public and private institutions in the respective member countries. The presentation, methods and questions of the Business Tendency Surveys are harmonised across countries according to EU guidelines. We refer to "The Joint Harmonised EU Programme of Business and Consumer Surveys User Guide 2002" of the European Commission, Directorate General Economic and Financial Affairs, Economic Studies and Research, Business Surveys for more detailed information on how these surveys are implemented. Two kinds of attitudinal data are collected: (i) "Judgments" (JUD), an assessment of the *current* or *past* status of a given variable, and (ii) "Expectations" (EXP), an estimation of the likely *future* status of that variable. These two types of attitudinal information are collected for five key components of the economy: Industry, Construction, Consumers, Retail Trade and Services. In this study, we focus on an *industry*-related indicator, the Production Expectations ($PEXP$).

The sample size for each survey varies across countries according to their population size. Each month, almost 68,000 companies and 27,000 consumers across the European Union are currently surveyed. Not surprisingly, Business Surveys are both expensive and time-consuming. For example, the cost in 2003 for the European Commission of the Harmonised Joint EU Programme of Business and Consumer Surveys amounted to slightly over 4.5 million euros. This figure does not yet incorporate the costs incurred by the respective national institutions that carry out the actual surveys. The latter are estimated to be another 6 to 7 million euros. These costs are often justified on the basis that they represent an early source of information for politicians, economists, researchers, senior executives and/or media. Indeed, the survey data become publicly available on the last day of each month, whereas Account data (ACC) - i.e. the actual, objectively measured, levels of a variable - are generally published only two to four months later (Buffeteau and Mora, 2000). However, *timeliness* is not a sufficient justification for the continued use of such expensive surveys. They should also prove to have some *predictive content* regarding actual ACC data: if Business Tendency Surveys cannot be used as relevant predictors for the future state of the economy, there is no point using them for economic surveillance.

The *performance* of the $PEXP$ series will be measured by studying whether these can improve forecasts of future Production Accounts ($PACC$) that are exclusively based on the past of the $PACC$ series. More precisely, we will investigate whether the variance of the error in forecasting future values of the *PACC*, using an (optimal) forecast based on the observed values of both *PACC* and *PEXP* series is *strictly smaller* than the variance of the prediction error, using an (optimal) forecast *only* based on the observed values of the *PACC* series. If this is the case, we may conclude that the *PEXP* series Granger causes the *PACC* series (Granger, 1969). In this chapter, the "predictive content" always refers to this notion of Granger causality.

The predictive content of Business Surveys has already been investigated in a substantive number of studies. Most of these were based on Scandinavian countries. Bergström (1995) and Teräsvirta (1986) explored the relationship between the *PACC* series and a number of barometer (Business Survey) series using regression models in, respectively, Sweden and Finland. Lindström (2000) extended these analyses by examining the short-term forecasting value of Business Surveys for predicting the volume of manufacturing output growth in Sweden up to 1998. Christoffersson et al. (1992) used a frequency domain analysis to investigate the relationship between *PACC* and Business Survey series of total manufacturing in Sweden. Öller and Tallbom (1996) applied an extended Kalman filter on the Business Survey series. The general picture obtained from these earlier studies is that Business Survey variables have predictive content in explaining future *PACC* series, at least in the considered Scandinavian countries. Hanssens and Vanden Abeele (1987), on the other hand, studied the predictive content of *PEXP* for Belgium, France, Germany, The Netherlands and Italy. Using bivariate Granger-causality tests, they found that the inclusion of *PEXP* did not systematically improve the forecasting performance of simple univariate time-series extrapolations of the *PACC* series. Hence, previous research ended in mixed results, and the question of the predictive value of Business Tendency Surveys remains open. We will contribute to this ongoing debate in a number of ways.

First, the existing literature focused on a limited subset of countries, ranging from one (e.g. Bergström, 1995 and Lindström, 2000) to five (Hanssens and Vanden Abeele, 1987) states. Our analysis, in contrast, uses data of more than ten countries (see the Data section). Moreover, by using information that has just become publicly available, we are able to extend the end date of the analysis from, respectively, 1983 (Teräsvirta, 1986; Hanssens and Vanden Abeele, 1987) and 1998 (Lindström, 2000) to December 2002.

Second, unlike previous studies, we will take potential *cross-country* relationships into account. Previous studies only allowed for a within-country relationship, checking, for example, whether the French *PEXP* proved useful in predicting France's own future *PACC* series. We extend the scope of this investigation by testing whether also the *PEXP* in *other* countries have information value in predicting the future evolution of the French *PACC* series. Such cross-country influences may well exist among the various EU countries as the free transfer of ideas, products and technologies among them is increasingly facilitated (Mahajan and Muller, 1994), resulting in both cross-country word-of-mouth effects (Tellis et al, 2003) and considerable trade among the various member states, as reflected in an intra-EU15 trade/GDP ratio of over 16% (Eurostat, 1999). Because of this economic interdependence, industrial optimism/pessimism in one country, as reflected in the evolution of its *PEXP*, may well affect the *PACC* in another country.

The aforementioned cross-country effects could be investigated through a sequence of *bivariate* tests. For example, bivariate Granger-causality tests could be implemented to assess whether *PEXP* in country X affects future *PACC* levels in country Y. However, as the number of EU countries increases, the number of tests required to cover all possible combinations would rapidly become excessive, causing the individual p-values to suffer from the well-known multiple-testing problem. As a third contribution, we will therefore apply the *multivariate* test procedure introduced by El Himdi and Roy (1996). Through this procedure, we will first investigate the general predictive content of the collection of all *PEXP* series vis-à-vis the collection of all *PACC* series. This joint Granger-causality test will offer a more powerful test on the predictive value of the ambitious European Union's Business Survey plan than a sequence of bivariate tests.

Next, we will adjust the El Himdi - Roy test statistic to answer three key questions:

- Can we attribute part of this "general causality" to *cross*-country influences?

- Are there some countries whose *PEXP* are more informative than others for the future evolution of the *PACC* in other countries? Such countries will be denoted as having a lot of "clout". Referring to the notion of asymmetry developed in the market-share literature (Cooper and Nakanishi, 1998, pp.56-57), "clout" is the ability of a brand or company to influence the other players (i.e. brands or firms) in a specific market, and consequently shape the demand and competition in that market. Similarly, in the context of *PEXP*, we denote by "clout" the ability of a country to Granger cause the future production level of other countries.

- Are there countries whose *PACC* series is more predictable (than others) by other countries' *PEXP* series, and whose economy therefore seems to be more "receptive" (see Cooper and Nakanishi (1998, pp.56-57) for a conceptually similar use of the term in the market-share attraction literature) to changes in other countries' economic climate (as reflected in their respective *PEXP* series)?

Answering these questions should provide new insights into the economic interdependencies among the various member states of the European Union in general, and more specifically, on the relative forecasting value of the Business Tendency Surveys undertaken in these countries.

3.2 Methodology

In line with existing literature (see e.g. Hanssens and Vanden Abeele, 1987), we adopt the notion of Granger causality between *PEXP* and future *PACC* to assess the predictive value of the Business Tendency Surveys' *PEXP* series. Unlike previous studies, however, we will conduct the analysis at various levels. First, we describe the bivariate Haugh (1976) test. The Haugh test will make it feasible to determine (i) whether the *PEXP* series in country i ($i = 1, ..., d$ where d is the number of considered countries) Granger causes the *PACC* series in that same country, and (ii) whether *PEXP* in country i Granger causes the *PACC* in country j. Next, we review the multivariate El Himdi - Roy (1997) test so as to allow to quantify the *combined* predictive content of all *PEXP* series, considering all countries together. We subsequently modify the El Himdi - Roy statistic to determine whether part of this combined predictive value can be attributed to *cross*-country influences. Finally, we propose two further modifications to quantify (i) which countries' *PEXP* series are the most informative as to the future *PACC* levels in other countries, i.e. have most "clout", and (ii) which countries' *PACC* series are most predictable through other countries' *PEXP*, i.e. have a higher "receptivity".

We denote *PACC* in country i at time t by Y_{it}, and use X_{it} to denote the surveyed *PEXP* made at t about the future production level in country i. Each of the aforementioned tests requires stationarity of the series under investigation. In line with prior studies (see e.g. Hanssens and Vanden Abeele, 1987; Öller and Tallbom, 1996), it was found, by carrying out various (unreported) stationarity tests, that all *PEXP* series (X_{it}) were already stationary, while the *PACC* series (Y_{it}) required seasonal differencing of order 12 ($\nabla^{12} Y_{it} = Y_{it} - Y_{it-12}$) to achieve stationarity.

3.2.1 Causality Tests based on Bivariate Cross-Correlations

For the bivariate test in country i, both time series, X_{it} and $\nabla^{12}Y_{it}$, are first modelled as a univariate ARMA process, i.e.

$$\begin{aligned} \Phi^x(L)X_{it} &= C^x + \Theta^x(L)u_{it} \\ \Phi^y(L)\nabla^{12}Y_{it} &= C^y + \Theta^y(L)v_{it} \end{aligned} \tag{3.1}$$

where $\Phi^x(L)$ and $\Phi^y(L)$ are autoregressive polynomials, $\Theta^x(L)$ and $\Theta^y(L)$ moving-average polynomials, and C^x and C^y the deterministic components (here, constant terms). After filtering the series with the above ARMA models, we obtain estimated innovation series $\hat{u}_{it}$ and $\hat{v}_{it}$, which can be regarded as white-noise processes with zero mean, but possibly correlated with each other at different leads and lags.

If we consider both series of innovations, u_{it} and v_{it}, it is well known (see e.g. Gouriéroux and Monfort, 1990, pp.368) that X_{it} is not Granger causing Y_{it} if and only if all cross-correlations at positive lags between u_{it} and v_{it} are equal to zero. So, with $\rho_{v_iu_i}(k) = corr(u_{it}, v_{it+k})$:

$$H_0 : X_{it} \nRightarrow Y_{it} \Leftrightarrow \rho_{v_iu_i}(k) = 0 \text{ for } k = 1, 2, 3, ..., \tag{3.2}$$

where $\nRightarrow$ means "does not Granger cause". Note that one could also test for the presence of feedback relationships between X_{it} and Y_{it} by considering the cross-correlations at *negative* lags between u_{it} and v_{it} instead (see e.g. Sims, 1972).

The bivariate cross-correlations are estimated as

$$\hat{\rho}_{v_iu_i}(k) = \frac{\sum_{t=1}^{T-k} \hat{u}_{it}\hat{v}_{it+k}}{\left(\sum_{t=1}^{T} \hat{u}_{it}^2\right)^{\frac{1}{2}} \left(\sum_{t=1}^{T} \hat{v}_{it}^2\right)^{\frac{1}{2}}}. \tag{3.3}$$

If the null hypothesis of no Granger causality holds, $\hat{\rho}_{v_iu_i}(k)$ should be close to zero for all positive lags. A test statistic for the hypothesis of no

Granger causality is then provided by Haugh (1976):

$$Q_M = T \sum_{k=1}^{M} \left(\hat{\rho}_{v_i u_i}(k)\right)^2 \backsim \chi^2_M \ , \tag{3.4}$$

where χ^2_M is a Chi-square distribution with M degrees of freedom. For a pre-specified value of M (here, the square root of the length of the series), Q_M is computed and the hypothesis of no Granger causality is rejected if $Q_M > \chi^2_{M,1-\alpha}$, where $\chi^2_{M,1-\alpha}$ is the $(1-\alpha)$ quantile of the Chi-square distribution with M degrees of freedom.

3.2.2 Multivariate Granger-Causality Tests

Let X_t and Y_t be two multivariate time series, $X_t \in \mathbb{R}^{d_1}$ and $Y_t \in \mathbb{R}^{d_2}$. In our case, $d = d_1 = d_2$, the number of EU countries under consideration. In a first step, the multivariate time series X_t and $\nabla^{12} Y_t$ are modeled separately, using Vector Autoregressive (VAR) models, i.e.

$$\begin{aligned} X_t &= C^x + \Phi_1^x X_{t-1} + \Phi_2^x X_{t-2} + \ldots + \Phi_{\rho_x}^x X_{t-\rho_x} + U_t \\ \nabla^{12} Y_t &= C^y + \Phi_1^y \nabla^{12} Y_{t-1} + \Phi_2^y \nabla^{12} Y_{t-2} + \ldots + \Phi_{\rho_y}^y \nabla^{12} Y_{t-\rho_y} + V_t, \end{aligned} \tag{3.5}$$

with U_t and V_t two, possibly correlated, multivariate white-noise series, $U_t \in \mathbb{R}^{d_1}$ and $V_t \in \mathbb{R}^{d_2}$. The order of each VAR model is determined by the SBIC (Schwartz Bayesian Information) Criterion. Furthermore, C^x and C^y are deterministic components; Φ_i^x and Φ_j^y are autoregressive polynomials, $\Phi_i^x \in \mathbb{R}^{d_1 \times d_1}$ for $i = 1, ..., \rho_x$ and $\Phi_j^y \in \mathbb{R}^{d_2 \times d_2}$ for $j = 1, ..., \rho_y$. We denote:

$$U_t = \begin{pmatrix} u_{1,t} \\ \vdots \\ u_{d_1,t} \end{pmatrix} \text{ and } V_t = \begin{pmatrix} v_{1,t} \\ \vdots \\ v_{d_2,t} \end{pmatrix}. \tag{3.6}$$

Note that the series U_t obtained after filtering with a VAR model are likely to differ from the series obtained by filtering each individual component of X_t through univariate models. Indeed, the innovations U_t obtained from (3.5) are independent of the past of every single component of X_t. With a univariate filtering procedure, in contrast, the innovation series from the first series may still carry information related to the past of the other series of X_t. Such indirect effects among the components of the X_t (or $\nabla^{12} Y_t$) vector are filtered out in the VAR-based approach of (3.5).

The estimated innovations U_t and V_t are cross-correlated at various leads and lags:

$$\begin{aligned} \hat{R}^{VU}(k) &= \mathrm{CORR}(\hat{U}_t, \hat{V}_{t+k}) \\ &= \begin{pmatrix} \hat{\rho}_{v_1 u_1}(k) & \hat{\rho}_{v_1 u_2}(k) & \dots & \hat{\rho}_{v_1 u_{d_1}}(k) \\ \vdots & & & \vdots \\ \hat{\rho}_{v_{d_2} u_1}(k) & \dots & \dots & \hat{\rho}_{v_{d_2} u_{d_1}}(k) \end{pmatrix} \in \mathbb{R}^{d_2 \times d_1} \end{aligned} \tag{3.7}$$

with $k = \dots, -2, -1, 0, 1, 2, \dots$

The corresponding auto-correlations are defined as

$$\begin{aligned} \hat{R}^{UU}(k) &= \mathrm{CORR}(\hat{U}_t, \hat{U}_{t+k}) \in \mathbb{R}^{d_1 \times d_1}, \\ \hat{R}^{VV}(k) &= \mathrm{CORR}(\hat{V}_t, \hat{V}_{t+k}) \in \mathbb{R}^{d_2 \times d_2}. \end{aligned} \tag{3.8}$$

Since the two series of innovations U_t and V_t are multivariate white noise, we know that $R^{UU}(k) = 0 = R^{VV}(k)$ for any $k \neq 0$.

Similar to (3.2), one can prove (see Gouriéroux and Monfort, 1990, pp.369-370) that

$$H_0 : X_t \nRightarrow Y_t \Leftrightarrow R^{VU}(k) = 0, \text{ for all } k > 0. \tag{3.9}$$

A test-statistic for the hypothesis of no Granger causality was proposed by El Himdi and Roy (1997). This test looks, for every $k = 1, \dots, M$ where M equals the maximum lag (again, the square root of the length of the series) at the vector:

$$\mathrm{vec}\, \hat{R}^{VU}(k) = \begin{pmatrix} \hat{\rho}_{v_1 u_1}(k) & \hat{\rho}_{v_2 u_1}(k) & ... & \hat{\rho}_{v_1 u_2}(k) & ... & \hat{\rho}_{v_{d_2} u_{d_1}} \end{pmatrix}^t \tag{3.10}$$

i.e. the correlation matrix $\hat{R}^{VU}(k)$ is vectorised into a vector of length $d_1 d_2$. Under the null hypothesis, all components of this vector should be small. Note though, that not all components are independent of each other. Indeed, one can show that, under the null hypothesis and for $k > 0$,

$$\operatorname{cov}(\hat{\rho}^{(k)}_{v_i u_j}, \hat{\rho}^{(k)}_{v_{i'} u_{j'}}) \approx \frac{R^{vv}(0)_{ii'} R^{uu}(0)_{jj'}}{T}, \tag{3.11}$$

with $1 \leq i, i' \leq d_2$ and $1 \leq j, j' \leq d_1$ and $\approx$ standing for asymptotical equivalence.

Using this vector, one can compute (similar to (3.4)) the following multivariate test-statistic:

$$Q^k_{HR} = T \left[\operatorname{vec} \hat{R}^{VU}(k) \right]^t A^{-1} \left[\operatorname{vec} \hat{R}^{VU}(k) \right], \tag{3.12}$$

where A is the asymptotic covariance matrix of $\sqrt{T} \operatorname{vec} \hat{R}^{VU}(k)$, so $A = \hat{R}^{VV}(0) \otimes \hat{R}^{UU}(0)$, with $\otimes$ the Kronecker product. The values of Q^k_{HR} can be plotted against k, giving an idea about the strength of the cross-correlation at various lags. The above quadratic form follows a $\chi^2_{d_1 d_2}$ distribution under the null hypothesis, leading to the El Himdi - Roy (1997) test-statistic:

$$Q_{HR} = \sum_{k=1}^{M} Q^k_{HR} \sim \chi^2_{M d_1 d_2}. \tag{3.13}$$

This multivariate procedure provides a more powerful test than the bivariate analogues previously described. The power gain is derived from two sources. First, all countries are pooled to find evidence of Granger causality. Moreover, it automatically permits to look for Granger causality across countries, meaning that we already allow for the possibility that the *PEXP* in one country Granger causes the *PACC* in another country. El Himdi and Roy (1997) applied this multivariate test to investigate the causal relations between money (M1 and M2) and income (Gross National Product) for Canada, as well as to study the causal directions between the Canadian and American economies.

Knowing the predictive content of *PEXP* in all considered EU countries together, one could wonder whether part of this causality can be attributed to cross-country influences. This is a natural extension of the bivariate cross-country analysis (see Equation 3.2). To test whether the joint cross-country causality is significant, we state the following null hypothesis:

$$H_0 : \tilde{R}^{VU}(k) = 0 \text{ for all } k > 0, \tag{3.14}$$

where $\tilde{R}^{VU}(k)$ is equal to $R^{VU}(k)$ without the elements on the diagonal, yielding a $d_1 \times (d_1 - 1)$ matrix.

The test statistic is adjusted accordingly to:

$$Q_{HR}^{cross} = T \sum_{k=1}^{M} \left[\text{vec}\, \widehat{\tilde{R}}^{VU}(k) \right]^t \tilde{A}^{-1} \left[\text{vec}\, \widehat{\tilde{R}}^{VU}(k) \right], \tag{3.15}$$

which is $\chi^2_{Md_1(d_1-1)}$ distributed, where $\tilde{A}$ is the asymptotic covariance matrix of $\sqrt{T}\,\text{vec}\,\widehat{\tilde{R}}^{VU}(k)$, $\tilde{A} = \widehat{\tilde{R}}^{VV}(0) \otimes \widehat{\tilde{R}}^{UU}(0)$. In fact, $\tilde{A}$ is simply equal to A, from which the rows and columns corresponding to the diagonal elements of $R^{VU}(k)$ are deleted. An analogous reasoning can be applied to obtain a within-country test statistic Q_{HR}^{within}, which will be $\chi^2_{Md_1}$ distributed.

Finally, in order to more deeply understand the underlying interdependencies within the European Union, one could further modify the multivariate procedure to also investigate whether the *PEXP* series in country i Granger causes the $PACC$ of all other countries $j = 1,..., d$, $i \neq j$, and therefore offer a quantification for that country's "clout". To test for the "clout" of country i, only the single component X_{it} has to be taken into account to test whether the latter Granger causes the collection of all other series Y_{jt}, with $j \neq i$. The El Himdi - Roy procedure is then applied as outlined above, with $d_1 = 1$ and $d_2 = d - 1$. Similarly, we may want to test whether the $PACC$ series of a given country j is Granger caused by the *PEXP* of all other countries $i = 1,..., d$, $i \neq j$, giving a measure of the "receptivity" of country j. In this case, $d_1 = d - 1$ and $d_2 = 1$. In both cases, the distribution of the test statistic is a $\chi^2_{(d-1)M}$.

3.3 Data

The Production Expectations *PEXP* series are provided by the Directorate General Economy and Finance of the European Commission.[1] They are by definition subjective, reflecting the respondents' optimism/pessimism w.r.t. the evolution of the production (excluding construction). They are expressed in Balance ($Bal = Pos - Neg$). Specifically, one asks the responding firms *whether they expect their production to increase, decrease or remain unchanged over the next three months*, and subsequently subtracts all *decrease* (Neg) answers - in percentage points of total answers - from the percentage of *increase* (Pos) ones. A directional questionnaire is used as directions of change have been found to be easier to predict than point values (Jonung, 1986). Balance data were also used in Bergström (1995) and Lindström (2000), among others.[2] Note, however, that problems with balance data may arise when positive and negative answers are not symmetrically distributed around zero (Entorf, 1993; Öller, 1990).

The Production Account *PACC* series are the monthly *Industrial Production Indices* reported in the OECD's publication, *Main Economic Indicators*.[3] These production indices are collected by the OECD from the different national institutions and are expressed as an index with 1995 scaled as base index 1 (at constant prices). Industrial production includes mining, manufacturing and production of electricity, gas and water, excluding construction (compatible with ISIC Rev.3).

We selected twelve EU countries out of the (before 2004) member states: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, The Netherlands and the United Kingdom.

[1] The data are publicly available on http://europa.eu.int/comm/economy_finance/indicators/businessandconsumersurveys_en.htm

[2] For an in-depth discussion on the relative merits of such balance indicators, see e.g. Granger (1980, ch.7), Klein and Moore (1983) or Hanssens and Vanden Abeele (1987).

[3] We refer to http://www.oecd.org/std/mei for more details. Note that the *PACC* series differ from the so-called National Accounts, which typically are quarterly data.

Three countries (Portugal, Spain and Sweden) were not withheld, as this would have resulted in the loss of multiple data points due to missing observations, since monthly surveys in these countries began later. However, the twelve retained EU countries represent over 85% of the Gross Domestic Product of the EU15 in 2002 (at current prices and exchange rates). All time series are collected on a monthly basis. The data range from January 1985 to December 2002, resulting in 216 monthly observations. As, among other remarkable events, Germany's reunification took place during the considered time span, we also examined the existence of structural breaks in the time series. According to standard CUSUM tests based on recursive least squares, structural stability was found to be plausible during the whole time period, and no structural dummies were included.

A visual inspection of the raw time series (see Figure 3.1) reveals some apparent similarities in the *PACC* series among various member countries. Several of the series show a pronounced upward trend, but suffered from a temporary slowdown in the early 1990s (see e.g. the graphs for Belgium, Finland, France, Germany and the United Kingdom, among others). Still, this slowdown was much less pronounced in countries like Denmark and The Netherlands, and even less so in Ireland. Note that the Irish *PACC* exhibit an curvilinear trend.[4]

Also in the *PEXP* series (see Figure 3.2), some similarities seem to be present (see in this respect the graphs for Belgium, France and Germany), while countries like Ireland and Denmark seem to follow a more idiosyncratic evolution.

In sum, this preliminary visual inspection suggests that (i) some cross-country commonalities and influences could well be present, but that (ii) this will probably not be homogeneous across all member states.

[4] A log-transform of the Irish series was considered, and found not affecting the results.

Figure 3.1: *Production Account PACC (raw) time series of the twelve EU countries.*

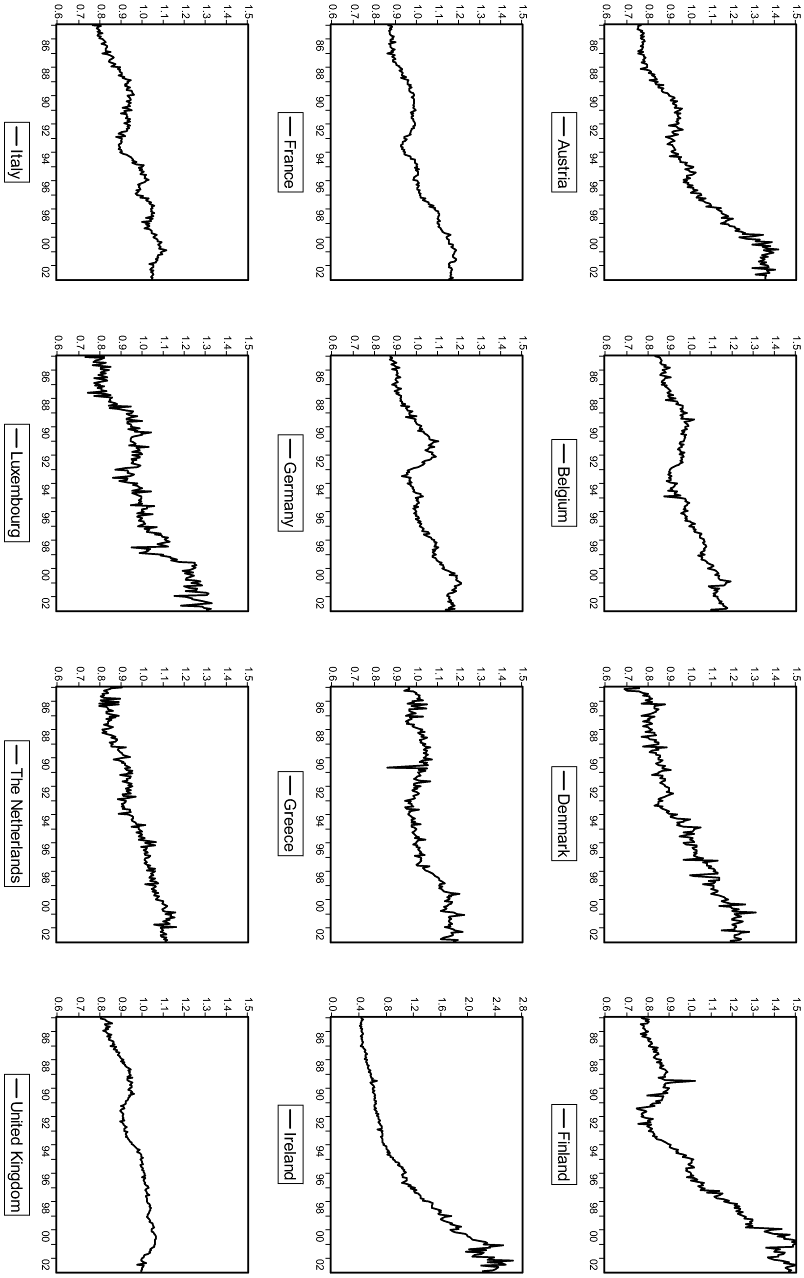
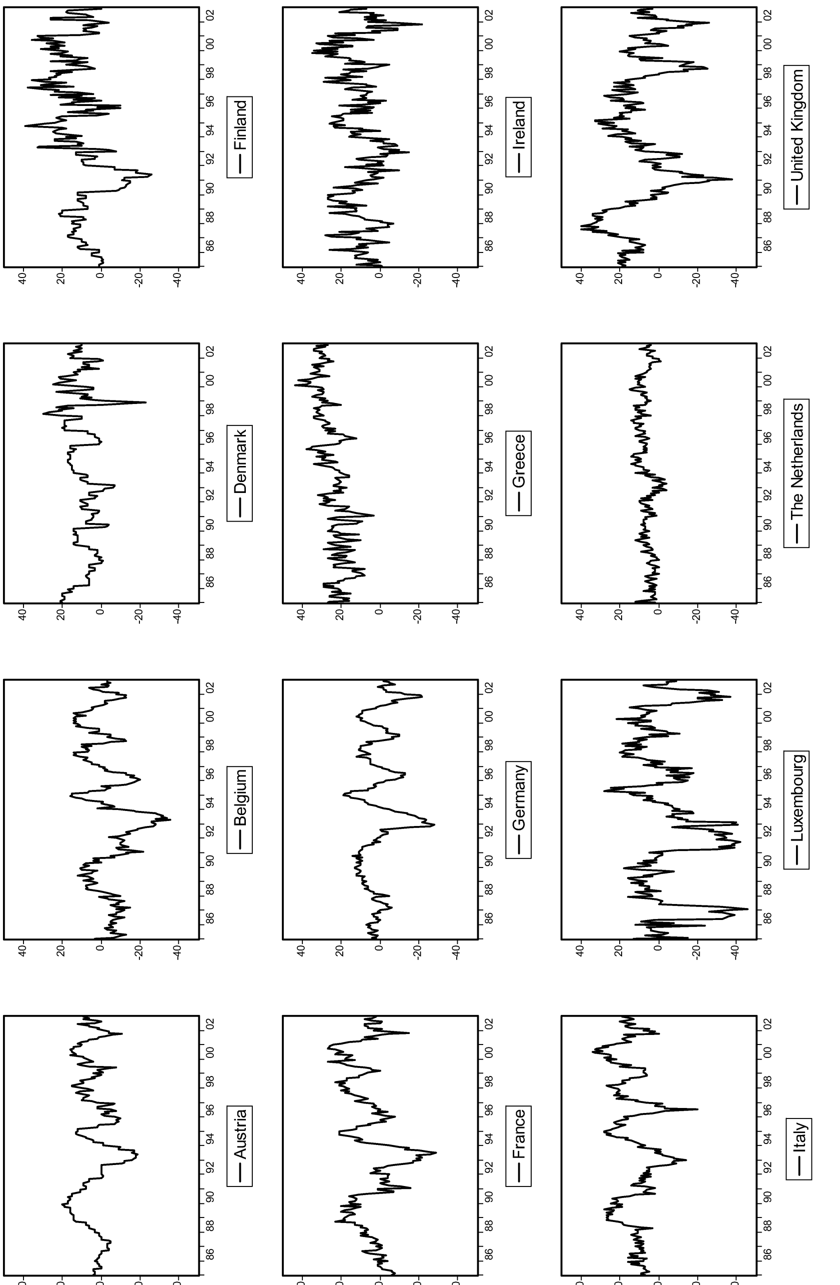


Figure 3.2: *Production Expectations PEXP (raw) time series of the twelve EU countries.*



3.4 Results

3.4.1 Bivariate Within-Country Predictive Content in Europe

Traditionally, the *within*-country forecasting value of the *PEXP* series was assessed by bivariate causality tests. This results, in our setting, in twelve tests, whose p-values are reported in the diagonal of Table 3.1.

The findings confirm our earlier observation that the evidence is mixed concerning the *within*-country predictive content of *PEXP*, even though some strong Granger causalities clearly appear in the table. Indeed, the null hypothesis of no Granger causality is rejected for 7 of the 12 countries at the 5% level. In line with Teräsvirta (1986), we find a significant predictive content for the Finnish *PEXP* series ($p = 0.000$). Note, however, that this result cannot be generalized to all Nordic countries, as the null hypothesis could not be rejected for Denmark ($p = 0.193$). This was also the case for Ireland ($p = 0.189$) and the United Kingdom ($p = 0.132$).

3.4.2 Bivariate Cross-Country Predictive Content in Europe

The possibility that *PEXP* may influence not only the own *PACC* series, but also (some) other countries' can be formally investigated by running an additional 132 bivariate Haugh tests, as summarized in the off-diagonal elements of Table 3.1. A value situated in the i^{th} row and j^{th} column of Table 3.1 represents the p-value for the test that the *PEXP* in country i Granger causes *PACC* in country j. Note that, because of the multiple-testing issue explained below, we test at a stricter significance level of 1%.

More than predicting its own *PACC*, German *PEXP* also Granger causes *PACC* changes in several other countries, i.e. in Belgium, Finland, France and The Netherlands (at the 1% level). A similar "clout" can

Table 3.1: *Bivariate cross-country analysis: p-values for the Haugh test for testing whether PEXP in country i (i^{th} row) Granger causes PACC in country j (j^{th} column).*

		Production Accounts *PACC*											
	p-value*	*Austria*	*Belgium*	*Denmark*	*Finland*	*France*	*Germany*	*Greece*	*Ireland*	*Italy*	*Luxembourg*	*The Netherlands*	*The United Kingdom*
Production Expectations *PEXP*	*Austria*	0.008*	0.246	0.019	0.234	0.002*	0.072	0.983	0.797	0.046	0.721	0.227	0.302
	Belgium	0.013	0.003*	0.003*	0.102	0.000*	0.137	0.068	0.052	0.006*	0.010*	0.019	0.015
	Denmark	0.004*	0.134	0.193	0.003*	0.185	0.780	0.927	0.247	0.139	0.530	0.311	0.191
	Finland	0.086	0.138	0.014	0.000*	0.048	0.293	0.694	0.012	0.268	0.103	0.510	0.125
	France	0.000*	0.002*	0.001*	0.013	0.000*	0.001*	0.008*	0.013	0.001*	0.051	0.002*	0.134
	Germany	0.125	0.000*	0.152	0.010*	0.003*	0.001*	0.036	0.552	0.162	0.039	0.006*	0.441
	Greece	0.395	0.016	0.152	0.023	0.393	0.107	0.080	0.448	0.183	0.297	0.272	0.056
	Ireland	0.032	0.053	0.039	0.290	0.395	0.105	0.961	0.189	0.460	0.019	0.042	0.486
	Italy	0.022	0.076	0.090	0.524	0.013	0.463	0.196	0.444	0.001*	0.009*	0.024	0.252
	Luxembourg	0.252	0.140	0.285	0.062	0.084	0.377	0.009*	0.709	0.200	0.013	0.347	0.931
	The Netherlands	0.131	0.020	0.138	0.021	0.293	0.071	0.145	0.674	0.027	0.444	0.057	0.311
	The United Kingdom	0.347	0.063	0.620	0.328	0.756	0.520	0.757	0.144	0.143	0.138	0.592	0.132

* significant at the 1% probability level

be attributed to France and Belgium. The former result is not surprising, as France and Germany are often seen as two of the key forces (both economically and politically) of the European unification. In addition, their respective populations are found to have among the highest rates of contact with other countries (Putsis et al., 1997). As for Germany, our findings provide further evidence in support of the German leadership hypothesis, thereby complementing the findings of Sensier et al. (2004), who found that Germany's composite leading indicator had a particularly marked effect on the Italian economy, and of Artis and Zhang (1998) and Barassi et al. (2000), among others, who studied the leading role of German interest rates. The finding for Belgium, in turn, may be linked to its central role in the location of the European Union's administration.

The British *PEXP*, in contrast, does not Granger cause any other country's *PACC* at the 1% probability level, not even its own. Even if this result cannot be overstated, it could reflect the distinct position of the United Kingdom in terms of geography, economy and culture (see Northcott, 1995 for a more elaborate discussion).

Note further that Table 3.1 is not symmetric. Consider, for example, Luxembourg. Its *PEXP* series is not informative for the future evolution of the *PACC* series of its neighboring Belgium, while that country's *PEXP* series Granger causes the *PACC* series of Luxembourg.

Even though the aforementioned examples provide some face validity to the figures of Table 3.1, it is obvious that the interpretation of that many test results is quite cumbersome, and may, not surprisingly, lead to some apparent "anomalies". For example, it is not immediately clear why *PEXP* in Luxembourg would Granger cause *PACC* in Greece, but not in its neighbouring, and culturally and economically more similar, Belgium. Moreover, even though 20 of the 132 off-diagonal p-values turned out to be smaller than 0.01, one should keep in mind that the p-values in Table 3.1 may suffer from the multiple-testing problem, and be biased downwards (Bauer et al., 1988) . Therefore, *the predictive power of the PEXP surveys is likely to be lower than what might be inferred from Table 1.* A multivariate extension of the bivariate Granger-causality tests can address these concerns.

3.4.3 Multivariate European Predictive Content

The El Himdi - Roy (HR) test for the null hypothesis of no Granger causality between *PEXP* and *PACC* at the joint European level resulted in a test statistic $Q_{HR} = 2687.760$, with associated p-value = 0.000. The multivariate test procedure therefore established that, in combination, the twelve *PEXP* series from the European Business Surveys' program have predictive value in forecasting the actual *PACC* series, thereby offering further justification for their continued use.

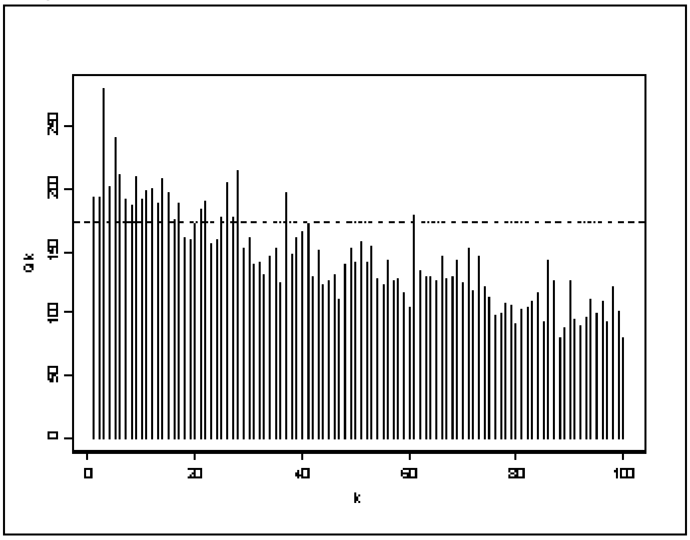
In order to get insight into the distribution of the predictive content of *PEXP* over time, we also consider the strength of the cross-correlation at lag k as measured by Q^k_{HR}, the k^{th} term of the test statistic Q_{HR}, defined in (3.12). In Figure 3.3, the values of Q^k_{HR} are plotted with respect to the lag length, and compared with the marginal critical value $\chi^2_{d_1 d_2, 1-\alpha}$, as in El Himdi and Roy (1997).

Figure 3.3 illustrates that the predictive content of the *PEXP* slowly decreases over time, and completely drops under the significance cut-off after a few years.

3.4.4 Multivariate Cross-Country Predictive Content in Europe

When testing for the joint nullity of all cross-correlations at positive lags (3.14), we obtain $Q^{cross}_{HR} = 2390.276$ (with $132 \times M$ degrees of freedom), with associated p-value = 0.000, and $Q^{within}_{HR} = 230.793$ (with $12 \times M$ degrees of freedom), with associated p-value = 0.000. This implies that cross-country influences clearly exist, and contribute to the high combined

Figure 3.3: *Measure* Q_{HR}^{k} *of multivariate Granger causality of PEXP on PACC, for lags up to* $k = 100$ *months.*



predictive content of *PEXP* series in Europe.

3.4.5 Multivariate Clout and Receptivity in Europe

Table 3.2 reports the results of the tests assessing the "clout" and the "receptivity" of each of the twelve EU countries. The second column provides the p-value for testing the Granger causality of the *PEXP* in the country mentioned in the first column on all other countries' *PACC* series. The smaller this value, the higher the "clout" of the country. The third column of Table 3.2 represents the p-value related to the test for Granger causality of the *PEXP* in all other countries on the *PACC* in that specific country. The smaller this value, the higher the "receptivity" of the country.

First of all, we may observe that only the *PEXP* series of Germany and France have highly significant predictive content (in the multivariate Granger sense) on future *PACC*. These findings are in line with our previous bivariate analysis, and confirm common intuition that those two economies are two key economic *drivers of the European integration.* The

Table 3.2: *p-values of the El Himdi - Roy test for assessing the "clout" and the "receptivity" of a selected country.*

Country	Clout	Receptivity
Austria	0.889	0.005*
Belgium	0.058	0.159
Denmark	0.810	0.029*
Finland	0.122	0.028*
France	0.000*	0.052
Germany	0.004*	0.427
Greece	0.096	0.253
Ireland	0.345	0.476
Italy	0.812	0.250
Luxembourg	0.098	0.148
The Netherlands	0.156	0.357
The United Kingdom	0.533	0.796

* significant at the 5% probability level.

economic climate in those countries (as reflected in their balance scores on the *PEXP* series) is found to Granger cause the subsequent actual production in the rest of the European Union, which offers further justification for the quite large sample sizes[5] (and hence higher costs of data collection) in those countries' Business Tendency Surveys. In contrast, Germany has a low "receptivity" score: while it drives the European Union, it is less likely to be influenced by the other member countries.

Another key fact illustrated by the multivariate test is that many countries have no significant "clout", not even at very liberal significance levels; see e.g. Austria ($p = 0.889$), Denmark ($p = 0.810$), Ireland ($p = 0.345$),

[5]According to "The Joint Harmonised EU Programme of Business and Consumer Surveys User Guide 2002", France and Germany together represent more than 25% of the industry-related surveyed units across the European Union.

Italy ($p = 0.812$), and the United Kingdom ($p = 0.533$). This suggests that the highly significant multivariate predictive content of the *PEXP* series will be mainly driven by a few key countries such as France and Germany. One could even question the usefulness of a continued use of the Business Tendency Surveys in countries like Denmark, Ireland or the United Kingdom, as they do not have a significant predictive power for the actual economic evolution in the other member countries (Table 3.2), nor for their own actual economic evolution (cf. diagonal elements of Table 3.1). The latter conclusion is, however, conditional on the analyzed *PEXP* and *PACC* series. As indicated in the Introduction section, the Business Tendency Surveys also collect data on other components of the economy such as retail trade, services, construction..., and more research is necessary to assess whether this conclusion also holds in these other domains.

In terms of the "receptivity", we find a significant ($p < 0.05$) effect for three countries: Austria, Denmark and Finland. Our findings for Denmark are in line with Putsis et al. (1997), who showed that this country derives a higher percentage of its contacts from outside its borders than internal. This corroborates the fact that Denmark has low within-country predictive power ($p = 0.193$), while having high "receptivity" ($p = 0.029$; Table 3.2). As for Austria and Finland, more than 60% of the total trade of these former EFTA countries is directed towards the EU (Sapir, 1998), which may well explain their high receptivity scores.

Finally, it is worth noting that the United Kingdom and Ireland score low on both dimensions. This may reflect their geographically distinct position, but also - at least for the United Kingdom - the country's relative distinctiveness in terms of economic integration and culture (Northcott, 1995). The low "receptivity" and "clout" of Ireland, in turn, is in line with the pan-European study of Mahajan and Muller (1994), who found the Irish population to be much less sensitive to word-of-mouth influences, which is a key mechanism through which cross-country *PEXP* may influence future *PACC* levels.

3.5 Conclusions

Each month, for over forty years, the European Union, together with institutions from the fifteen member states of the European Union, carries out costly and time-consuming Business Tendency Surveys on the past, actual and future state of the European Economy. Since they are, on a regular basis, used for economic surveillance by several parties, there is a concern about their *performance* in predicting actually realized Account data, especially since existing literature on the topic offers mixed evidence on the Surveys' predictive content.

In this chapter, we tested the Granger causality between *PEXP* and *PACC* series from twelve European countries. Instead of only focusing on a sequence of bivariate within-country analyses, we also undertook a simultaneous, *multivariate* approach. Indeed, when predicting the future production level in a country, there is no reason to a priori ignore the information conveyed by Business Surveys in other countries. In this sense, the predictive content of *PEXP* series could be evaluated both at the *national* (bivariate test) and at the *European* (multivariate test) level.

While confirmation of Granger causality at the individual country level was not found for every country - seven countries out of twelve showed a significant predictive content at the 5% level - , very strong evidence of Granger causality was discovered at the multi-country level. It also turned out that the cross-country Granger causalities were jointly strongly significant, indicating that it could indeed be advantageous to exploit these correlations in multivariate forecast methods.

Moreover, according to the multivariate analysis, some countries (i.e. France and Germany) have more "clout" than others, while others are more "receptive" (i.e. Austria, Finland and Denmark). Finally, the United

Kingdom and Ireland seem to occupy a fairly isolated position. They both have no “clout” and no “receptivity”.

As a conclusion, we state that the harmonization of European Business Tendency Surveys allows to exploit cross-country relations between the different series and to improve forecasts of future Account data for an individual country by using a multivariate approach. More research is open to generalize these findings to other variables that are routinely collected in the EU Business Tendency Surveys.

References

[1] Artis, M.J. and Zhang, W. (1998). The Linkage of Interest Rates within the EMS, *Weltwirtschaftliches Archiv*,134, 117-132.

[2] Barassi, M.R.,Caporale, G.M. and Hall, S.G. (2000). Interest Rate Linkages: A Kalman Filter Approach to Detecting Structural Change, *Imperial College Management School Working Papers*, 5 (June).

[3] Bauer, P., Hommel, G. and Sonnemann, E. (1988). *Multiple hypotheses testing*, Berlin: Springer.

[4] Bergström, R. (1995). The Relationship between Manufacturing Production and Different Business Survey Series in Sweden 1968-1992, *International Journal of Forecasting*, 11, 379-393.

[5] Buffeteau, S. and Mora, V. (2000). Predicting the National Accounts of the Euro Zone using Business Surveys, *INSEE Conjuncture in France*, December.

[6] Christoffersson, A., Roberts, R. and Eriksson U. (1992). The Relationship between Manufacturing and Various Business Tendency Survey Series in Sweden Illuminated by Frequency and Complex Demodulate Methods, *NIER Working Paper Stockholm*, 15.

[7] Cooper, L.G. and Nakanishi, N. (1998). *Market-Share Analysis*, Massachusetts: ISQM Kluwer Academic Publishers.

[8] El Himdi, K. and Roy, R. (1997). Tests for Non-Correlation of two Multivariate ARMA Time Series, *The Canadian Journal of Statistics*, 25, 233-256.

[9] Entorf, H. (1993). Constructing Leading Indicators from non-Balanced Sectoral Business Surveys Series, *International Journal of Forecasting*, 9, 211-225.

[10] Gouriéroux, Ch. and Monfort, A. (1996). *Time Series and Dynamic Models*, Glasgow: Cambridge University Press.

[11] Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, 37, 424-438.

[12] Granger, C.W.J. (1980). *Forecasting in Business and Economics*, New York: Academic Press.

[13] Hanssens, D.M. and Vanden Abeele, P.M. (1987). A Time-Series Study of the Formation and Predictive Performance of EEC Production Survey Expectations, *Journal of Business and Economic Statistics*, 5, 507-519.

[14] Haugh, L.D. (1976). Checking the Independence of two Covariance-Stationary Time Series: a Univariate Residual Cross-Correlation Approach, *Journal of the American Statistical Association*, 71, 378-385.

[15] Jonung, L. (1986). Uncertainty about Inflationary Perceptions and Expectations, *Journal of Economic Psychology*, 7, 315-325.

[16] Klein, P.A. and Moore, G.H. (1983). The Leading Indicator Approach to Economic Forecasting - Retrospect and Prospect, *Journal of Forecasting*, 2, 119-135.

[17] Lindström, T. (2000). Qualitative Survey Responses and Production over the Business Cycle, *Central Bank of Sweden - Working Paper Series*, 116, November.

[18] Mahajan, V. and Muller, E. (1994). Innovation Diffusion in a Borderless Global Market: Will the 1992 Unification of the European Community Accelerate Diffusion of New Ideas, Products, and Technologies?, *Technological Forecasting and Social Change*, 45, 221-235.

[19] Northcott, J. (1995). The Future of Britain and Europe, *London Policy Studies Institute Research Report,* 776.

[20] Öller, L.-E. (1990). Forecasting the Business Cycle Using Survey Data, *International Journal of Forecasting*, 6, 453-461.

[21] Öller, L.-E. and Tallbom, C. (1996). Smooth and Timely Business Cycle Indicators for Noisy Swedish Data, *International Journal of Forecasting*, 12, 389-402.

[22] Putsis, W.P., Balasubramanian, S., Kaplan E.H. and Sen, S.K. (1997). Mixing Behavior in Cross-Country Diffusion, *Marketing Science*, 16, 354-369.

[23] Sapir, A. (1998). The Political Economy of EC Regionalism, *European Economic Review*, 42, 717-732.

[24] Sensier, M., Artis, M., Osborn, D.R. and Birchenhall, C. (2004). Domestic and International Influences on Business Cycles Regimes in Europe, *International Journal of Forecasting*, 20, 343-357.

[25] Sims, C.A. (1972). Money, Income, and Causality, *The American Economic Review*, 62, 540-552.

[26] Tellis, G.J., Stremersch, S. and Yin E. (2003). The International Takeoff of New Products: The Role of Economics, Culture and Country Innovativeness, *Marketing Science*, 22, 188-208.

[27] Teräsvirta, T. (1986). Model Selection with Business Survey Data: Forecasting the Output of Finnish Metal and Engineering Industries, *International Journal of Forecasting*, 2, 191-200.

Chapter 4

Decomposing Granger Causality over the Spectrum

4.1 Introduction

Investigating causality is a topic of main interest in scientific research. To assess the causality between two processes in a common and well-defined (non-experimental) framework, one usually refers to the well-known concept of *Granger* causality (GC), introduced in 1969 by the 2003 Nobel prize winner in Economics. GC reflects a restricted sense of causality, i.e. *the extent to which a process X_t is leading another process Y_t*, and builds upon the notion of incremental predictability. Specifically, a process X_t Granger causes another process Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t. The reader should keep in mind that, in some circumstances, the aforementioned notion of causality may not fully coincide with the concept of causation (see [1]). However, the underlying intuition for this approach is that if an event is the cause of another, it should precede it. Therefore, as soon as the application of the causality concept refers to the search for the best predictive model, the concept of GC is valid. The standard test of GC developed by Granger [2] is based on the following regression model

$$Y_t = \alpha_0 + \sum_{k=1}^{M} \beta_{1k} Y_{t-k} + \sum_{k=1}^{M} \beta_{2k} X_{t-k} + \varepsilon_t, \qquad t = M+1, ..., T \quad (4.1)$$

where ε_t are uncorrelated random variables with mean zero and variance σ^2, and M is the specified lag length. The null hypothesis that X_t does not Granger cause Y_t is supported when $\beta_{2k} = 0$ for $k = 1, ..., M$, causing (4.1) to reduce to

$$Y_t = \alpha_0 + \sum_{k=1}^{M} \beta_{1k} Y_{t-k} + \tilde{\varepsilon}_t. \quad (4.2)$$

A wide range of bivariate GC tests exist,[1] which have been used extensively to study a wide range of substantive, economic, issues. For instance, the well-known "export-led growth" hypothesis has been studied repeatedly in a GC framework (see e.g. [5]), as has been the relationship between economic growth and various other variables, such as business-cycle volatility [6], the degree of openness [7], and defense spending [8]. In the financial literature, GC testing has been applied, for instance, to identify price-leadership patterns among national stock prices [9], to study the stock price-volume relationship [10], to get insight into the dynamic behavior of bonds and stocks [11], or into the causal relationship between equity and real estate returns [12]. In marketing, GC testing has been used predominantly to discern competitive reactions patterns (see Hanssens *et*

[1]One of them is the well-known Granger-Wald test (see [2]). It is defined as

$$GW = T \frac{(\hat{\sigma}^2_{\tilde{\varepsilon}_t} - \hat{\sigma}^2_{\varepsilon_t})}{\hat{\sigma}^2_{\varepsilon_t}}$$

where $\hat{\sigma}^2_{\tilde{\varepsilon}_t}$ is an estimate of the variance of $\tilde{\varepsilon}_t$ from model (2), and $\hat{\sigma}^2_{\varepsilon_t}$ is an estimate of the variance of ε_t from model (1), and follows an asymptotic χ^2_M distribution under the null hypothesis. Another well-established GC test is the double prewhitening technique known as the Haugh-Pierce [3] test (see next section for more details). Both tests have been found to have a reasonable power together with a small bias in size (type-I error) (see e.g. [4] for a review).

al. [13], p.314 for a review). While not exhaustive, the above enumeration clearly demonstrates the widespread use of the GC concept in both economics and business.

In this chapter, we propose a spectral-density based GC test. This approach offers some distinct advantages relative to the aforementioned, standard, procedures. While traditional tests indicate whether Granger causality is present or not, we propose to give a richer and more complete picture by decomposing Granger causality over different frequencies or time horizons. As such, one can, for example, compare the predictive power present at the short, middle or long run. The spectral GC test can be applied at any given frequency superior to the sampling frequency of the data (in our application, one month). This allows us to gain insights into potential variations in the strength of the GC between the two variables over the spectrum. Indeed, there is increasing research evidence that the nature of the relationship between two variables may vary depending on the time horizon under consideration. Such variation was, for example, found in the relationship between real exchange rates and real interest differentials [14], between the GDP series of different countries [15], in the nature of competitive price reactions [16], in the effects of price promotions [17], and in the link between aggregate advertising spending and various macro-economic indicators [18]. Baxter [14], for example, found evidence of a relationship between real exchange rates and real interest differentials at trend (long-run) and business-cycle (middle-run) frequencies (i.e. low to middle frequencies), which was not found in prior studies that only focused on high-frequency components. Similarly, Croux *et al.* [15] found that the GDPs of US states are more correlated with each other than European countries' together, but also that this difference is much more pronounced in the short run (i.e. at the high frequencies). As a consequence, a one-shot GC test that is supposed to apply across all time horizons (e.g. in the short run, over the business cycle frequencies, and in the long run) may well give an incomplete, and potentially misleading, picture of the temporal ordering between the variables of interest.

Measures and tests for Granger causality in the frequency domain have already been proposed by Geweke [19] and Hosoya [20], and have recently been reconsidered in Breitung and Candelon [21]. Our test for GC is similar in spirit, but uses a nonparametric estimate of the cross-spectrum between filtered time series. An important advantage of the procedure proposed in this chapter is that the tests for GC at different frequencies are asymptotically independent of each other. This makes it possible to easily carry out a joint test for GC at different frequencies of the spectrum simultaneously. More details are provided in the next section.

Note that the cross-correlogram between Y_t and X_t is not an appropriate tool to decompose the GC over different time horizons. First of all, the estimated cross-correlations between X_t and Y_t are strongly correlated. Moreover, cross-correlations are very cumbersome to interpret, since there is confoundedness between correlations *within* the series and *between* the series (see e.g. [22], p. 139). Hence, we propose to use the value of the spectral-based GC test, and plot these measures with respect to the frequency, resulting in a graphical tool which gives insight into the decomposition of the GC over the spectrum. This plot is essentially given by the Fourier transform of the cross-correlations at negative lags between filtered versions of the series X_t and Y_t.

The remainder of this chapter is structured as follows. The spectral GC test is detailed in the next section. Subsequently, we apply this test to investigate the predictive value of European production expectation surveys. Finally, we offer some brief conclusions.

4.2 A Spectral Granger-Causality Approach

Let X_t and Y_t be stationary (after possible transformations) time series. Spectral analysis is performed on the innovations series, u_t and v_t, derived from X_t and Y_t. The latter are modeled as univariate ARMA processes, i.e.

$$\begin{aligned}\Phi^x(L)X_t &= C^x + \Theta^x(L)u_t \\ \Phi^y(L)Y_t &= C^y + \Theta^y(L)v_t\end{aligned} \tag{4.3}$$

where $\Phi^x(L)$ and $\Phi^y(L)$ are autoregressive polynomials, $\Theta^x(L)$ and $\Theta^y(L)$ moving-average polynomials, and C^x and C^y potential deterministic components. After filtering the series with the above ARMA models, we obtain the innovation series u_t and v_t, which are white-noise processes with zero mean, possibly correlated with each other at different leads and lags. These innovations are central to the development of the well-known Haugh-Pierce [3] test for GC,[2] and will also form the main building blocks for our proposed testing procedure.

Let $S_u(\lambda)$ and $S_v(\lambda)$ be the spectral density functions, or spectra, of u_t and v_t at frequency $\lambda \,\epsilon\, [-\pi, \pi]$ defined by

$$S_u(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_u(k)e^{-i\lambda k} \quad \text{and} \quad S_v(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_v(k)e^{-i\lambda k}, \tag{4.4}$$

where $\gamma_u(k) = Cov(u_t, u_{t-k})$ and $\gamma_v(k) = Cov(v_t, v_{t-k})$ represents the autocovariances of u_t and v_t at lag k. The idea of the spectral representation is that each time series may be decomposed into a sum (or integral) of uncorrelated components, each related to a particular frequency λ. A detailed treatment on the spectral analysis of time series is given in Koopmans [23] or Warner [24].

[2] Under the null hypothesis of no GC, the M cross-correlations $\rho_{vu}(k) = corr(v_t, u_{t-k})$ between the innovation series u_t and v_t, with $k = 1, ..., M$, are asymptotically independently and normally distributed with mean zero and standard deviation $T^{-1/2}$, and the Haugh-Pierce [3] test statistic

$$HP = T\sum_{k=1}^{M}\hat{\rho}^2_{vu}(k)$$

is asymptotically chi-square distributed with M degrees of freedom.

As the innovations series u_t and v_t are white-noise processes, the spectra (4.4) are constant functions, given by

$$S_u(\lambda) = \frac{Var(u_t)}{2\pi} \quad \text{and} \quad S_v(\lambda) = \frac{Var(v_t)}{2\pi}.$$

Therefore, their spectra $S_u(\lambda)$ and $S_u(\lambda)$ can simply be estimated as

$$\hat{S}_u(\lambda) = \frac{\widehat{Var}(u_t)}{2\pi} \quad \text{and} \quad \hat{S}_v(\lambda) = \frac{\widehat{Var}(v_t)}{2\pi}.$$

To investigate the relationship between both stochastic processes under consideration, we consider the cross-spectrum, $S_{uv}(\lambda)$, between u_t and v_t. This is a complex number, defined as

$$S_{uv}(\lambda) = C_{uv}(\lambda) + iQ_{uv}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{uv}(k) e^{-i\lambda k}, \tag{4.5}$$

where the cospectrum $C_{uv}(\lambda)$ and the quadrature spectrum $Q_{uv}(\lambda)$ are, respectively, the real part and the imaginary part of the cross-spectrum. Here $\gamma_{uv}(k) = Cov(u_t, v_{t-k})$ represents the cross-covariance of u_t and v_t at lag k. The cross-spectrum can be estimated non-parametrically by

$$\hat{S}_{uv}(\lambda) = \frac{1}{2\pi} \left\{ \sum_{k=-M}^{M} w_k \hat{\gamma}_{uv}(k) e^{-i\lambda k} \right\}, \tag{4.6}$$

with $\hat{\gamma}_{uv}(k) = \widehat{Cov}(u_t, v_{t-k})$, the empirical cross-covariances, and with window weights w_k, for $k = -M, \ldots, M$. The expression (4.6) is called the *weighted covariance estimator*, and when the weights w_k are selected as $1 - \frac{|k|}{M}$, the Barlett weighting scheme is obtained. The constant M is the maximum lag order considered.[3]

This cross-spectrum allows to compute the coefficient of coherence $h_{uv}(\lambda)$, defined as (see [23])

$$h_{uv}(\lambda) = \frac{|S_{uv}(\lambda)|}{\sqrt{S_u(\lambda) S_v(\lambda)}}. \tag{4.7}$$

[3]In practice (see e.g. [25], p.136), M is often chosen to be equal to the square root of the number of observations T.

This coefficient, which can take on values between zero and one, gives a symmetric measure of the strength of linear association between two time series, frequency by frequency, but does not express any information on the *direction* of the relationship between two processes. The squared coefficient of coherence, which has a similar interpretation as the R-squared in a regression context, was used in Barksdale *et al.* [26] to study the association between advertising and sales, or in Woitek [27] to investigate the relationship between human height cycles and cycles of economic variables.

A confidence interval for the coefficient of coherence can be derived. Specifically, for $n > 20$, we have that $\tanh^{-1}(h_{uv}(\lambda))$ is well approximated by a normal distribution:

$$\tanh^{-1}\left(\hat{h}_{uv}(\lambda)\right) \approx N\left(\tanh^{-1}(h_{uv}(\lambda)), \frac{1}{2(n-1)}\right), \qquad (4.8)$$

with $n = T/\sum w_k^2$ (see e.g. [23]). We denote $2(n-1)$ as the equivalent degrees of freedom, EDF. From the above equation, if follows immediately that one can reject the null hypothesis of the nullity of the coefficient of coherence if

$$|\hat{h}_{uv}(\lambda)| > \tanh\left(z_{\alpha/2}\sqrt{\frac{1}{EDF}}\right),$$

where $z_{\alpha/2}$ is the $\frac{\alpha}{2}$ upper quantile of a standard normal distribution.

If one wants to take the *direction* of the relationship into account, as asked for to indicate GC, the coefficient of coherence in (4.7) should be adapted. Specifically, the cross-spectrum (4.5) can be decomposed into three parts, (i) $S_{u\Leftrightarrow v}$, the instantaneous relationship between u_t and v_t, (ii) $S_{u\Rightarrow v}$, the directional relationship between v_t and lagged values of u_t, and (iii) $S_{v\Rightarrow u}$, the directional relationship between u_t and lagged values of v_t, i.e.

$$\begin{aligned} S_{uv}(\lambda) &= \frac{1}{2\pi}\left[S_{u\Leftrightarrow v} + S_{u\Rightarrow v} + S_{v\Rightarrow u}\right] \qquad (4.9)\\ &= \frac{1}{2\pi}\left[\gamma_{uv}(0) + \sum_{k=-\infty}^{-1}\gamma_{uv}(k)e^{-i\lambda k} + \sum_{k=1}^{\infty}\gamma_{uv}(k)e^{-i\lambda k}\right]. \end{aligned}$$

The proposed spectral measure of GC is based on the key property that X_t does not Granger cause Y_t if and only if $\gamma_{uv}(k) = 0$ for all $k < 0$ (see [28]). Hence, if the goal is to assess the predictive content of X_t relative to Y_t, one is mainly interested in the second part of (4.9), i.e.

$$S_{u \Rightarrow v}(\lambda) = \frac{1}{2\pi}\left[\sum_{k=-\infty}^{-1} \gamma_{uv}(k) e^{-i\lambda k}\right].$$

A Granger coefficient of coherence is then given by

$$h_{u \Rightarrow v}(\lambda) = \frac{|S_{u \Rightarrow v}(\lambda)|}{\sqrt{S_u(\lambda) S_v(\lambda)}}. \tag{4.10}$$

Therefore, in the absence of GC, given that the numerator in (4.10) cancels out, $h_{u \Rightarrow v}(\lambda) = 0$ for every λ in $[-\pi, \pi]$. A natural estimator for the Granger coefficient of coherence at frequency λ is

$$\hat{h}_{u \Rightarrow v}(\lambda) = \frac{\left|\hat{S}_{u \Rightarrow v}(\lambda)\right|}{\sqrt{\hat{S}_u(\lambda) \hat{S}_v(\lambda)}},$$

with $\hat{S}_{u \Rightarrow v}(\lambda)$ as in (4.6), but with all weights w_k for $k \geq 0$ put equal to zero. Similarly as for the coefficient of coherence distribution (4.8), one has that the transformed GC coefficient is approximately normally distributed[4]

$$\tanh^{-1}\left(\hat{h}_{u \Rightarrow v}(\lambda)\right) \approx N\left(\tanh^{-1}\left(h_{u \Rightarrow v}(\lambda)\right), \frac{1}{2(n'-1)}\right) \tag{4.11}$$

with

$$n' = \frac{T}{\sum_{k=-M}^{-1} w_k^2}.$$

Indeed, since the weights w_k with a positive index k are set equal to zero when computing $\hat{S}_{u \Rightarrow v}(\lambda)$, only the w_k with negative indices need to be taken into account when computing the appropriate degrees of freedom

[4]The weights in (4.6) are usually taken to be symmetric, in the sense that $w_{-k} = w_k$. Allowing for asymmetric weights, however, does not alter the asymptotic results, as it can be seen by verifying the proofs in Brillinger [29].

$EDF' = 2(n'-1)$. The null hypothesis of no Granger causality at frequency λ, formally $H_0 : h_{u \Rightarrow v}(\lambda) = 0$, is then rejected if

$$|\hat{h}_{u \Rightarrow v}(\lambda)| > \tanh\left(z_{\alpha/2}\sqrt{\frac{1}{EDF'}}\right). \qquad (4.12)$$

Our spectral-based GC approach is very flexible as it provides a measure of GC at each individual frequency λ of choice, making it feasible to investigate the strength of the causal relationships, e.g. at the short run, at the business cycle frequency, or at the long run. Relative to the other existing GC tests (cf. footnote 1), the spectral-based approach therefore provides additional information on the frequency where GC is (most) prevalent, or, in contrast, negligible.

As mentioned in the introduction, other spectral-based GC tests do exist (see e.g. [28], p.373-382). Based on the earlier work of Geweke [19] and Hosoya [20], Breitung and Candelon [21] have recently constructed a test procedure by imposing a set of linear restrictions on the autoregressive parameters in a bivariate Vector Autoregressive (VAR) framework. They apply this test to investigate the predictive content of the yield spread for future output growth. Our approach differs from earlier spectral-based GC tests in a number of ways. First, it does not require a bivariate VAR specification, but uses a non-parametric estimation of the cross-spectrum. Second, it is closely related to the well-established measure of (squared) coherence (see e.g. [30]), the equivalent of the R-squared in the frequency domain, making the measure of GC proposed here easier to interpret. As a third difference, the spectral estimates $\hat{h}_{u \Rightarrow v}(\lambda)$ used in our framework are asymptotically independent of each other at different frequencies, which is not the case for the previously proposed tests. This independence property allows us to easily perform a joint test for a null hypothesis of the form $H_0 : h_{u \Rightarrow v}(\lambda_1) = \ldots = h_{u \Rightarrow v}(\lambda_s) = 0$, with $\lambda_1, \ldots, \lambda_s$ being different frequencies. Indeed, under H_0, one has

$$EDF \sum_{j=1}^{s} \left(\tanh^{-1}(\hat{h}_{u \Rightarrow v}(\lambda_j)) \right)^2 \approx \chi_s^2. \tag{4.13}$$

We illustrate the use of such a joint test in the next section.

4.3 Application to the European Production Surveys

4.3.1 Introduction

Various governments conduct, at regular intervals, a wide range of surveys about the judgments and anticipations of consumers, producers and/or manufacturers. As one of the largest-scale surveys, the European Union has collected, for over forty years, so-called Business Tendency Surveys. About 68,000 companies and 27,000 consumers across the European Union are surveyed each month about (i) their *judgments* (i.e. their assessment of the current or past status of a given variable), and (ii) their *expectations* (i.e. their estimation of the likely future status of that variable). Even though such surveys are costly to conduct, they may offer useful leading information on the underlying economic variables. Indeed, the actual, objectively measured, values of those variables, denoted as *accounts*, typically become available only several months later than the surveys' results [31]. However, timeliness may not be a sufficient condition to justify the high costs involved, and the accuracy of the Business Tendency Surveys should be another key consideration. Specifically, one should expect (or hope) that these attitudinal measures have good predictive power, a question that naturally translates into the notion of GC. If a surveyed variable Granger causes its complementary account variable, the respondents of the survey would be found to possess implicit knowledge about the future levels of the account variable, that could not be derived from previous account levels.

Even though various studies have already considered whether judgmental data reported in the European Union's Business Tendency Surveys Granger cause the corresponding objective measures (see, in this respect, [32] or [33]), none of them has decomposed this potential predictive power over different frequencies or time horizons. Using the procedure previously outlined, it is now possible to compare the strength of the GC for the short-run, middle-run and longer-run components of the series. We can test, for example, whether the strength of the causality at the short run is significant or not. If the former is verified, the Business Tendency Surveys are able to pick up the quickly-changing (i.e. short-run) components of the series. Our prior belief is that the slowly-moving (i.e. long-run) component would be more accurately forecasted by the surveys' respondents than the more unpredictable fast-moving changes in production. Testing whether the GC is significant at a long-run frequency boils down to check whether the surveys have predictive value for the more slowly-varying component of the corresponding account series.

4.3.2 Data

In this illustration, we focus on the (potential) predictive value of the European Production Expectations series. These publicly available data[5] are provided by the Directorate General Economy and Finance of the European Union. They reflect the respondents' optimism/pessimism w.r.t. the evolution of the production, and are expressed in Balance ($Bal = Pos - Neg$). Specifically, one asks the responding firms whether they expect certain variables to increase, decrease or remain stable over time, and subsequently subtracts all *decrease* (Neg) answers - in percentage points of total answers - from the percentage of *increase* (Pos) ones. This expectation series will form our X_{it} series (with $i = 1, ..., 12$), as data

[5]See http://europa.eu.int/ comm/economy_finance/indicators/ businessandconsumersurveys_en.htm

will be used on 12 countries (i.e. Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, The Netherlands and the United Kingdom).[6] The data range from January 1985 to December 2002, resulting in 216 observations.

The investigated account time series, i.e. the Y_{it} series for each country i, are the European Production Accounts series, which are published as part of the National Accounts Statistics by the OECD.[7] They are expressed as an index with 1995 scaled as base index 1 (at constant prices). These production account series have been used extensively (see e.g. [34]). All time series are collected on a monthly basis, and are already seasonally adjusted by the data providers. Note that, as Germany's unification took place during the considered time span, we also examined the existence of structural breaks in the time series. According to standard CUSUM tests based on recursive least squares, structural stability was found to be plausible during the whole time period, and no structural dummies were included. In the sequel, we formally investigate whether the European production expectations have significant predictive power, and if so, at what time frame, with respect to their corresponding future production accounts.[8]

4.3.3 Empirical Results

Since the traditional time-domain-based GC tests as well as the spectral GC tests require stationarity of time series, we seasonally differenced the production account series[9] (Y_{it}), as in Lemmens *et al.* [33]. In line with prior studies (see e.g. [32] or [35]), it was found, by carrying out var-

[6] Three countries (Portugal, Spain and Sweden) were not withheld, as this would have resulted in the loss of multiple data points due to missing observations, since surveys in these countries began later.

[7] See OECD publication, *Main Economics Indicators*, the Industrial Production Index, ref. 2027K

[8] For all undergone tests, we took $M = \sqrt{T}$

[9] The Irish account series was taken in logarithm due to its exponential trend

ious (unreported) stationarity tests, that all production expectation series (X_{it}) were already stationary. Next, the series were filtered[10] to obtain white-noise processes, which, as indicated in the previous section, form the building blocks for our Granger coefficient of coherence.

For each of the twelve countries, we present in Figure 4.1 the estimated Granger coefficients of coherence $\hat{h}_{u \Rightarrow v}(\lambda)$, for all $\lambda \in [0, \pi]$. This coefficient tests whether the production expectations are Granger causing the production accounts of a given country at that frequency. The baseline represents the critical bound of significance at the 5% probability level, as given in (4.12). Note that the frequency λ on the horizontal axis can be translated into a cycle of T months by $T = 2\pi/\lambda$ (in months, for monthly data).

Figure 4.1 shows a consistent pattern across eight countries, i.e. Austria, Belgium Finland, France, Germany, Italy, The Netherlands, and the United Kingdom. For those countries, the GC at small frequencies (corresponding to the longer-run components) clearly dominates those at the higher frequencies (corresponding to the shorter-run components). Hence, even though these countries' production expectations are found to have significant (incremental) predictive power with respect to the longer-run components in the production account series, they have much more difficulty in predicting the fast-moving components of these series. Apart from Germany, the Granger coefficients of coherence corresponding with the high frequencies hardly reach statistical significance, and also for Germany, the GC measure remains much more pronounced at the lower frequencies. However, given these countries' significant Granger coefficient of coherence at the lower part of the frequency band, we expect, for these countries, that also an overall GC test will indicate significance. As indicated in Table 4.1, this conjecture was confirmed through a formal Granger-Wald test in seven

[10] According to different diagnostic tests, residuals obtained after *SARIMA* modeling did not deviate significantly from white-noise processes (Figures are available upon request).

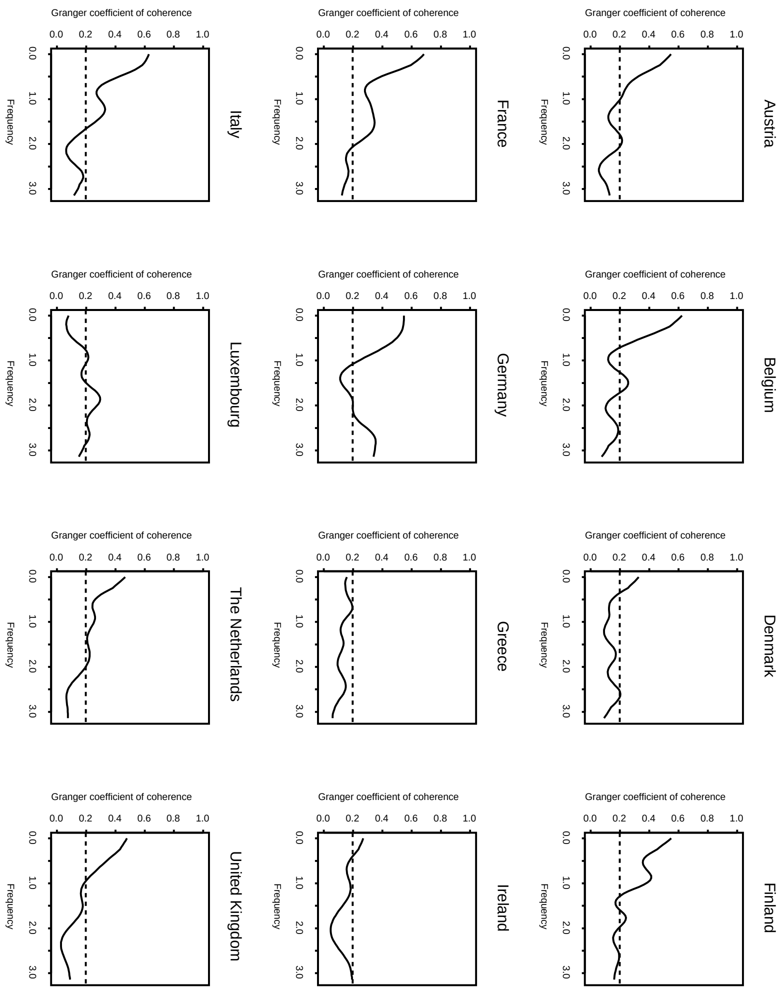


Figure 4.1: *Granger coefficients of coherence for 12 European countries. The dotted line represents the critical bound at the 5% probability level.*

instances ($p < 0.01$), while the test statistic for the remaining country (The Netherlands) was only marginally insignificant ($p = 0.056$). A similar picture emerged when applying the Haugh-Pierce test as "overall" GC test, even though the corresponding p-values were somewhat higher.[11] A possible explanation for this loss in power is that for these eight countries, the GC is concentrated with the low-frequency components (see Figure 4.1), while the Haugh-Pierce test does not give enough weight to this part of the spectrum.

Other countries, in contrast, do not show much evidence of GC between production expectations and accounts at any of the frequencies. This is the case for Greece, Ireland, and, to a lesser extent, Denmark and Luxembourg. As a consequence, for those countries, we do not expect an overall GC test to give a significant outcome either. Again using the conventional Granger-Wald test, we found support for this conjecture for Ireland ($p = 0.357$) and Denmark ($p = 0.146$). For Luxembourg, a significant overall test statistic was found ($p < 0.05$), which may be attributed to the significant GC coefficients of coherence between frequencies 1.5 and 2.5. This significant overall GC corresponds, however, with a very different coherence pattern than the one observed for the previous eight countries. This difference in the nature of the GC relationship over different time horizons goes undetected in conventional test procedures. When applying the Haugh-Pierce test, an insignificant effect was found for three countries (Denmark, Greece and Ireland), while, as with the Granger-Wald test, a significant overall GC was found for Luxembourg.

A few countries also have some idiosyncratic features. For example, we observe a small increase in the GC coherence measure around frequency 1.5 (corresponding with cycles of 4 months) in Belgium, France and Italy, while Finland experiences such a small increase around frequency 1.0. Rather than trying to explain each of these idiosyncratic features, we find it more

[11] This affected the outcome of the test statistic in only one instance, i.e. the United Kingdom.

insightful to focus on the general picture that emerges across the various countries. Specifically, the Granger coefficient of coherence tends to either decrease in λ (for eight out of twelve countries), or to remain flat and non-significant over the whole frequency band (see e.g. Greece, Denmark and Ireland). For the latter countries, the production expectations do not convey additional information about future production levels, which puts into question the usefulness of spending considerable amounts of money in collecting these data. Such additional information is conveyed for the other countries, however, but (except for Germany) *only* for the longer-run evolution of the production series. Hence, while production expectation series may be available earlier than the actual account series, this timeliness does *not* translate into an incremental forecasting ability for the fast-moving (high-frequency) movements in the account series. This misconception would not have been detected through conventional GC tests, and nicely illustrates the additional insights that can be obtained through the proposed spectral decomposition approach.

Table 4.1: *p-values for the Granger-Wald (G-W) test, Haugh-Pierce (H-P) test, as well as our coherence-based GC test, at three different frequencies for testing GC between production expectations and accounts (with $M = \sqrt{T}$). Significant p-values at the 5% probability level are reported in bold.*

p-value	Traditional tests		Spectral-based tests			
Country	G-W test	H-P test	$\lambda_1 = 0.5$	$\lambda_2 = 1.5$	$\lambda_3 = 2.5$	Joint test ($s = 3$)
Austria	**0.006**	**0.008**	**0.012**	0.400	0.946	0.069
Belgium	**0.000**	**0.003**	**0.005**	**0.007**	0.065	**0.000**
Finland	**0.000**	**0.000**	**0.007**	0.221	0.052	**0.006**
France	**0.000**	**0.000**	**0.003**	**0.001**	0.094	**0.000**
Germany	**0.000**	**0.001**	**0.000**	0.398	**0.004**	**0.000**
Italy	**0.000**	**0.001**	**0.000**	**0.008**	0.152	**0.000**
The Netherlands	0.056	0.057	0.071	0.054	0.616	0.065
United Kingdom	**0.002**	0.132	**0.003**	0.077	0.790	**0.008**
Denmark	0.146	0.193	0.453	0.196	**0.038**	0.089
Greece	**0.003**	0.080	0.067	0.146	0.120	**0.048**
Ireland	0.357	0.213	0.160	0.208	0.292	0.197
Luxembourg	**0.001**	**0.013**	0.284	0.130	**0.038**	0.051

The differential intensity of the GC relationship across different frequencies is also evident in Table 4.1, where we present the outcomes of formal spectral-based tests at three different frequencies: $\lambda_1 = 0.5$, $\lambda_2 = 1.5$ and $\lambda_3 = 2.5$. The choice of these frequencies is motivated by the timeliness of the surveys.[12] In line with our earlier discussion, we find (i) that considerable more GC is found at the low frequency (seven countries) than at the high frequency (three countries), and (ii) that for some countries (e.g. Ireland), there is little support for GC at any frequency, as opposed to countries like Belgium, France, Germany and Italy (among others) where there is significant GC at various frequencies. This differential behavior is also reflected in the final column of Table 4.1, which jointly considers the significance of the GC relationship across all three frequencies.

To conclude, the spectral-based GC test complements and extends the insights obtained through traditional CG tests. For example, from Table 4.1, we can infer that the highly significant overall GC tests for Finland, which are in line with earlier findings by Bergström [36] and Teräsvirta [37], are primarily due to the ability of the Finnish production expectation series to better predict the long-run evolution of their production accounts' counterparts. For Luxembourg, in contrast, this is due to frequencies in the 1.5-2.5 range. Also when comparing Germany and France, interesting new insights emerge. The Granger-Wald and Haugh-Pierce test both indicate, in line with Lemmens *et al.* [33], a highly significant predictive power for their respective production expectation series. The spectral tests, however, show that, for France, superior predictions are only obtained for periods longer than three months, while German production expectation series have significance predictive content at high frequencies as well.

[12] Recalling that the business tendency surveys are collected every month, and are mainly aimed at providing early information on the evolution of the production account series, relatively higher frequencies are expected to be more relevant than the lowest frequencies. Hence, we perform spectral-based tests for GC at frequencies corresponding to cycles of approximately 2.5 months, 4 months, and 12.5 months.

4.4 Conclusions

As indicated before, GC has been used extensively in previous work to study a wide range of substantive economic issues. Even though there is increasing evidence that the nature of the relationship may vary with the time horizon under consideration (see e.g. [14]), most previous studies have applied an overall GC test. In this chapter, we provided a new, spectral-based, approach that offers insights into potential variations in the strength of the GC over different frequencies. We demonstrated the additional insights that could be obtained with this testing procedure in the context of the forecasting ability of European-wide expectation surveys. We believe that comparable additional insights could also be obtained in several other substantive areas.

However, a number of interesting areas for future research remain. First, while we have shown how a joint test for GC can be carried out for a finite number of distinct frequencies, we did not yet develop a test procedure to test for the nullity of $h_{u \Rightarrow v}(\lambda)$ over a subinterval $\lambda \in [a, b]$ of $[-\pi, \pi]$. The distribution of such a test statistic would be very complicated, and is beyond the scope of the current chapter. As another limitation, we restricted ourselves to a test for bivariate GC. In some settings however (see e.g. [33]), it could be insightful to also carry out multivariate tests of GC.

References

[1] Zellner A. Causality and Econometrics. *Journal of Monetary Economics (Supplement)*. 1979; **10**: 1-54.

[2] Granger CWJ. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*. 1969; **37**: 424-438.

[3] Pierce DA, Haugh LD. Causality in Temporal Systems. *Journal of Econometrics*. 1977; **5**: 265-293.

[4] Bult JR, Leeflang PSH, Wittink DR. The Relative Performance of Bivariate Causality Tests in Small Samples. *European Journal of Operational Research*. 1997; **97**: 450-464.

[5] Abual-Foul B.Testing the Export-Led Growth Hypothesis: Evidence from Jordan. *Applied Economics Letters*. 2004; **11**: 393-396.

[6] Dopke J. How Robust is the Empirical Link between Business-Cycle Volatility and Long-Run Growth in OECD Countries? *International Review of Applied Economics*. 2004; **18**: 103-121.

[7] Bahmani-Oskooee M, Niroomand F. Openness and Economic Growth: An Empirical Investigation. *Applied Economics Letters*. 1999; **6**: 557-561.

[8] Al-Yousif YK. Defense Spending and Economic Growth: Some Empirical Evidence from the Arab Gulf region. *Defence Peace Economics*. 2002; **13**: 187-197.

[9] Peiers B. Informed Traders, Intervention, and Price Leadership: A Deeper View of the Microstructure of the Foreign Exchange Market. *Journal of Finance*. 1997; **52**: 1589-1614.

[10] Hemstra C, Jones JD.Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation. *Journal of Finance*. 1994; **49**: 1639-1664.

[11] Park J, Shenoy C. An Examination of the Dynamic Behavior of Aggregate Bond and Stock Issues. *International Review of Economics and Finance.* 2002; **11**: 175-189.

[12] Okunev J, Wilson P, Zurbruegg R. Relationships between Australian real estate and stock market prices: a case of market inefficiency. *Journal of Forecasting*, 2002; **21**: 181-192.

[13] Hanssens DM, Parsons LJ, Schultz RL. *Market Response Models: Econometric and Time Series Analysis*, 2nd Edition, Kluwer Academic Publishers: Amsterdam, 2001.

[14] Baxter M. Real Exchange Rates and Real Interest Differentials: Have We Missed the Business-Cycle Relationship? *Journal of Monetary Economics.* 1994; **33**: 5-38.

[15] Croux C, Forni M, Reichlin L. A Measure of Comovement for Economic Variables: Theory and Empirics. *The Review of Economics and Statistics.* 2001; **83**: 232-241.

[16] Bronnenberg BJ, Mela CF, Boulding W. The Periodicity of Competitor Pricing. *Working paper.* 2005; The Andersen School of Management, UCLA.

[17] Dekimpe MG, Hanssens DM, Nijs VR, Steenkamp J-BEM. Measuring short- and long-run promotional effectiveness on scanner data using persistence modelling. *Applied Stochastic Models in Business and Industry.* 2005; **21**, 4-5, 409-416.

[18] Deleersnyder B, Dekimpe MG, Leeflang PSH. Advertising and the Macro-Economy: Have we Missed the Business-Cycle Relationship? 2004; *Paper presented at the 33th EMAC Conference, Murcia 2004.*

[19] Geweke J. Measurement of Linear Dependence and Feedback between Multiple Time Series, *Journal of the American Statistical Association.* 1982; **77**: 304-324.

[20] Hosoya Y. The Decomposition and Measurement of the Interdependence between Second-Order Stationary Processes, *Probability Theory and Related Fields.* 1991; **88**: 429-444.

[21] Breitung J, Candelon B. Testing for Short- and Long-Run Causality: A Frequency-Domain Approach, *Journal of Econometrics. in press.*

[22] Chatfield C. *The Analysis of Time Series: An Introduction*, fifth edition, Chapman and Hall: London, 1996.

[23] Koopmans LH. *The Spectral Analysis of Time Series*, Probability and Mathematical Statistics 22, Academic Press: San Diego, 1995.

[24] Warner RM. *Spectral Analysis of Time-Series Data*. The Guilford Press: New York, 1998.

[25] Diebold FX. *Elements of Forecasting*, second edition, South-Western: Ohio, 2001.

[26] Barksdale HC, Hilliard JE, Guffey HJ. Spectral Analysis of Adverising-Sales Interaction: An Illustration. *Journal of Advertising*. 1974; **3**: 26-33.

[27] Woitek U. Height Cycles in the 18th and 19th Centuries. *Economics and Human Biology*. 2003; **1**: 243-257.

[28] Gouriéroux C, Monfort A, Gallo G. *Time Series and Dynamic Models*, Cambridge University Press: Glasgow, 1996.

[29] Brillinger DR. Asymptotic Properties of Spectral Estimates of Second Order. *Biometrika*. 1969; **56**: 375-390.

[30] Brockwell PJ, Davis RA. *Introduction to time series and forecasting*, second edition, Springer Texts in Statistics: New York, 2002.

[31] Buffeteau S, Mora V. Predicting the National Accounts of the Euro Zone using Business Surveys. *INSEE Conjuncture in France*. 2000; December.

[32] Hanssens DM, Vanden Abeele PM. A Time-Series Study of the Formation and Predictive Performance of EEC Production Survey Expectations. *Journal of Business and Economic Statistics*. 1987; **5**: 507-519.

[33] Lemmens A, Croux C, Dekimpe MG. The Predictive Content of Production Surveys: a Pan-European Study. *International Journal of Forecasting*, 2005; **21**: 363-375.

[34] Kawasaki Y, Franses PH. Do seasonal unit roots matter for forecasting monthly industrial production? *Journal of Forecasting*. 2004; **23**: 77-88

[35] Öller L-E, Tallbom C. Smooth and Timely Business Cycle Indicators for Noisy Swedish Data. *International Journal of Forecasting*. 1996; **12**: 389-402.

[36] Bergström R. The Relationship between Manufacturing Production and Different Business Survey Series in Sweden 1968-1992. *International Journal of Forecasting*. 1995; **11**: 379-393.

[37] Teräsvirta T. Model Selection with Business Survey Data: Forecasting the Output of Finnish Metal and Engineering Industries. *International Journal of Forecasting*. 1986; **2**: 191-200.

Chapter 5

The European Consumer: United in Diversity?

5.1 Introduction

Western-European countries have a longstanding post-WW II tradition of unification, as reflected in agreements to establish the Benelux, the European Free Trade Association, the European Union, and eventually, the European Monetary Union (Tellis, Stremersch, & Yin, 2003; see also McDonald & Dearden, 2005 for an extensive discussion). Also the increasing mobility, education and sophistication of consumers, the growing availability of various distance-spanning technologies, and the emergence of pan-European media have contributed to the perception that distance has become irrelevant within Europe (Mahajan & Muller, 1994; Tellis et al., 2003; Ter Hofstede, Steenkamp, & Wedel, 1999). All these factors suggest that the different member states could be treated as a single market, making a unified, pan-European marketing strategy appropriate (Steenkamp & Ter Hofstede, 2002). Such a strategy is attractive not only because of the economies of scale that European standardization may leverage (Yip, 1995), but also by the possibility to coordinate competitive and strategic moves, or to exploit the emergence of global retailers (Özsomer & Simonin, 2004). However, one could also argue that European countries continue to differ considerably from each other, economically (The Economist, 1999), in terms of laws and regulations (The European Voice, 2001), and (some may argue, especially) as far as cultural identity is concerned (de Mooij and Hofstede, 2002; Kraus, 2003; Rosenberger, 2004). If countries continue to have predominantly distinct market identities, multi-domestic, rather than pan-European, marketing strategies are called for.[1]

Previous research on the "unity" of the European market has provided mixed evidence. One stream of research supports pan-European marketing strategies. Ter Hofstede et al. (1999), for example, identify a pan-European consumer segment in yoghurt consumption patterns. In Gielens and Dekimpe (2001), neither cultural nor geographical proximity is found to affect the long-run performance of European retailers' international operations. In their study on the drivers of consumer acceptance of new packaged goods, Gielens and Steenkamp (2004) report that various consumer variables work in the same direction in four key European countries (France, Germany, Spain and the United Kingdom), suggesting that these variables offer a basis for horizontal market segmentation across borders.

Other studies, in contrast, identify substantial differences between various European countries, providing support for multi-domestic or multi-regional strategies. Geographic, economic and/or cultural distances are then found to remain key drivers of market heterogeneity in Europe. Bijmolt, Paas, and Vermunt (2004), for instance, find that European countries differ considerably in financial-product ownership. Based on that dimen-

[1] In many circumstances, a "compromise" strategy between standardization versus local flexibility is rather advocated (e.g. Boote, 1983; Holt, Quelch, & Taylor, 2004; Kamakura, Novak, Steenkamp, & Verhallen, 1994; Yip, 1995). This necessity of balancing is often referred to as "glocal" strategies. This idea is to globalize some parts of the strategy (e.g. production, organization, technologies), while customizing other parts according to the local and cultural specificities (e.g. product features, communication, distribution).

sion, they partition the European market in seven segments. Interestingly, their division is closely linked with geographical proximity. In terms of food culture, Askegaard and Madsen (1998) find Europe to be heterogeneous across its geographical and language borders. In the diffusion literature, Tellis et al. (2003) report substantially different times-to-takeoff for new products in Europe, partially related to cultural distances. Stremersch and Tellis (2004), in turn, discover significant differences in the European growth rates of consumer durables, and find these differences to be mainly related to economic distances. Finally, Kumar, Ganesh, and Echambadi (1998) conclude that geographical, economic and cultural distances help to explain diffusion similarities across Europe.

In sum, research on the unity of the European market offers mixed conclusions. One reason could be that all aforementioned studies consider *domain-specific* segmentation bases (Wedel & Kamakura, 1998), covering specific characteristics as yoghurt consumption, financial-product ownership, or takeoff of consumer durables. While such insights are very useful to the particular industry, they are less likely to generalize to other settings (Steenkamp & ter Hofstede, 2002). We therefore adopt a more *general* measure of consumer homogeneity/heterogeneity in Europe that is less dependent on the specific domain of study. Our point of departure is the Consumer Confidence Indicator (CCI) of the various European countries, which has been shown, in a wide variety of settings, to be a useful predictor of consumers' willingness to buy and future expenditures (see e.g. Nahuis & Jansen, 2004). Indeed, the European CCI and its US counterpart, the Index of Consumer Sentiment (ICS), have been found to be leading indicators of consumer expenditures on durables (Burch & Gordon, 1984; Throop, 1992), non-durables (Mueller, 1963), household goods and motor vehicles (Friend & Adams, 1964; Adams, 1965), and fashion merchandise (Allenby, Jen, & Leone, 1996), among others. In addition, they have been found to be useful in forecasting recession periods (Batchelor & Dua, 1998), and can be used as a proxy for consumer sunspots, i.e. changes of attitudes (Chauvet & Guo, 2003).

As a consequence, the CCI seems an obvious candidate to study in more general terms the extent of homogeneity in consumers' attitudes and buying behavior. The construct also offers some other advantages: these publicly available data are collected consistently by the European Commission over multiple countries and over a long time span. Moreover, as the construct is conceptually similar to the American ICS, a formal comparison with the United States, which has a much longer history of unification, becomes feasible.

As a second contribution, we analyze the degree of homogeneity in European consumers' CCI *dynamically.* Previous research is typically based on *static* similarity measures. Bijmolt et al. (2004), for example, partition the European market in terms of a one-shot measure of product ownership; ter Hofstede et al. (1999) segment means-end relations identified in a single data-collection wave; and also Askegaard and Madsen's (1998) analysis of European food cultures is based on lifestyle survey data collected at a single point in time. While international diffusion-based studies consider multiple data points, their main focus lies in subsequently explaining the cross-sectional variation in a single summary statistic, such as the time-to-takeoff (Tellis et al., 2003), average growth rate (Stremersch & Tellis, 2004), or asymptotic value (Gielens & Dekimpe, 2001). However, there is increasing evidence that the relationship between economic variables may vary, in direction and/or importance, over different planning horizons (see e.g. Baxter, 1994). In marketing, numerous studies have demonstrated that the short and long-run effectiveness of marketing-mix expenditures may differ considerably (see e.g. Nijs, Dekimpe, Steenkamp, & Hanssens, 2001; Pauwels, Hanssens, & Siddarth, 2002). Bronnenberg, Mela and Boulding (2004) find that the nature of competitive interactions differs (cooperative versus competitive) for different planning cycles, and Deleersnyder, Dekimpe, & Leeflang (2004) find that the link between aggregate advertising and GNP over business-cycle frequencies differs from relationships found in the short and long run. Indirect evidence for the relevance of this time dependence in assessing the usefulness of pan-European marketing strategies is provided in the combined studies of Tellis et al. (2003) and Stremersch and Tellis (2004). Using the same European diffusion data, they find different factors (respectively, cultural and economic) to drive the time-to-takeoff and subsequent growth rate of consumer durables. Hence, depending on the planning stage, different country segments emerged.

In this chapter, we study how the homogeneity in European CCIs varies as the planning horizon is extended. Indeed, country-specific disturbances may dampen the extent of short-run homogeneity, while more homogeneous patterns could come out as the planning horizon is extended. Should this be the case, the feasibility/attractiveness of pan-European marketing strategies will depend on the planning horizon one envisions. A myopic (short-run) focus may then inspire managers to adopt a multi-country strategy, foregoing the potential longer-run benefits of a pan-regional, or even pan-European, strategy.

To formally investigate this possibility, we apply the *dynamic-correlation* and *cohesion* concepts (Croux, Forni & Reichlin, 2001) to the evolution of the Consumer Confidence Indicators. In so doing, we address the following questions. First, to what extent are the CCIs homogeneous across all Member States of the European Union? How does this degree of homogeneity differ across different planning horizons, and how does it compare to the homogeneity across the different regions of the United States? Second, if there is considerable heterogeneity across the Member States, do certain regions (segments) exist which are more homogeneous? Finally, to what extent can geographic, cultural and economic distances help explain the observed heterogeneity, if any, in the various countries' CCI?

The remainder of the chapter is organized as follows. In Section 5.2, we formally discuss the concepts of dynamic correlation and cohesion, which are derived in the spectral domain. In Section 5.3, we discuss the data, and present empirical findings in Section 5.4. Managerial implications and conclusions are drawn in Section 5.5.

5.2 Dynamic Correlations

5.2.1 Spectral Analysis

Most currently available time-series applications in marketing are situated in the time domain (see Dekimpe & Hanssens, 2004 for a recent review). Spectral analysis, situated in the frequency domain and very popular in engineering (see e.g. Priestley, 1981), has received much less attention. Early exceptions are Parsons and Henry (1972), Barksdale, Hilliard and Guffey (1974), and Barksdale, Hilliard and Ahlund (1975). Parsons and Henry (1972) introduced spectral analysis as a diagnostic tool to test the equivalence between actual and predicted sales series. Barksdale et al. (1974) applied spectral tools to study the relationship between advertising expenditures, car factory sales, and new-car registrations over different frequencies. Finally, Barksdale et al. (1975) studied the link between price changes and quantities of beef at the slaughter level. Short-run changes in price were found to lead short-run changes in quantity by several months. In contrast, long-run decreases in quantities corresponded to long-run increases in price without time delay.

More recently, Bronnenberg et al. (2004) investigated the nature of competitive price reactions occurring at different frequencies. They found competitors' reactions to short-term price reductions to differ considerably from their reactions to long-run prices changes. In the former case, there was clear evidence of cooperative behavior between brands (i.e. the reactions are negatively correlated), while competitive behavior prevailed in the longer run (i.e. the correlation is positive). Finally, Deleersnyder, Dekimpe, Sarvary, and Parker (2004) used spectral band-pass filters in their study on the link between the durables' diffusion patterns and business-cycle fluctuations.

A common finding in the above studies is that marketing relationships may differ across different frequencies (planning horizons). This led Pauwels et al. (2005) to call for more spectral-based time-series applications in marketing, as this could lead to novel insights into a wide variety of substantive marketing problems.

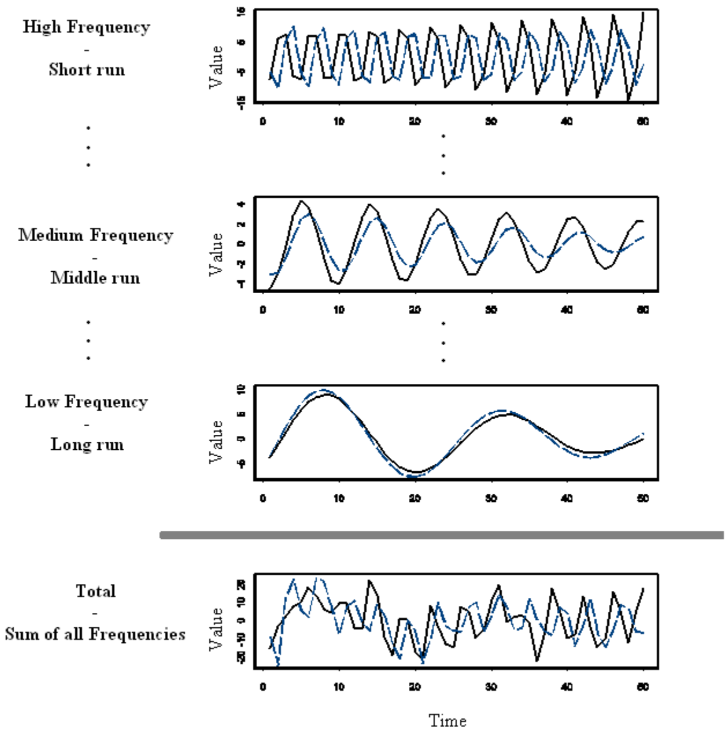
Central to spectral theory is the notion that any time series can be decomposed into an infinite sum of (uncorrelated) cyclical components, each having a different frequency λ. Each frequency λ (ranging between 0 and π) corresponds to a unique time horizon T, with $T = \frac{2\pi}{\lambda}$. In case of monthly data, a frequency of 0.5 represents a one-year time horizon (more precisely, 12.56 months), i.e. the yearly cyclical component in the time series. The underlying intuition is illustrated in Figure 5.1 for two simulated processes. Both series are formed by higher-frequency components (corresponding to shorter-run time horizons), middle-frequency components (for middle-run time horizons), and lower-frequency components (for longer-run time horizons). All components are added to each other to compose the time series, as illustrated in the final plot of Figure 5.1.[2]

In reality, a time series is composed of an infinite sum of such components, which can be isolated through spectral analysis. This makes it possible to study the correlation between two time series at any time horizon. In Figure 5.1, we see that the high-frequency components ($T \approx 4$ months) are quite uncorrelated, having different amplitudes and being out-of-phase.[3] On the contrary, the low-frequency components ($T \approx 24$

[2]The time series have been simulated over 50 months. The first plot represents a typical high-frequency component ($\lambda = 1.57$, $T \approx 4$ months), the second a typical medium-frequency component ($\lambda = 0.70$, $T \approx 12$ months), and the third a typical low-frequency component ($\lambda = 0.26$, $T \approx 24$ months). Components at other frequencies are not reported since an infinity of components forms the time series.

[3]The *amplitude* of the cyclical components is given by the height of the waves. Waves

Figure 5.1: *The decomposition of two time series in components at different frequencies.*



months) are almost perfectly correlated, as their amplitudes are very close and the series are in phase.

Let us now consider N stationary time series $x_1, \ldots, x_N$ of length T. In our application, the series represent the (first-differenced) CCI of the various EU countries. Traditional unit-root tests can be used to test for the stationarity of the various series (see e.g. Pauwels et al., 2002, or Nijs et al., 2001, for recent marketing applications). Removal of stochastic trends - by first differencing the series - is called for, as this trend would otherwise be treated as part of a very long oscillation, which would swamp the effects

having the same frequency but with their maxima occurring at different instances are said to be *out-of-phase*.

of shorter-period data (Parsons & Henry, 1972). Each stationary series x_i is characterized by a spectral density function, or spectrum $S_{x_i}(\lambda)$, which is defined at each frequency $\lambda \in [0, \pi]$ by

$$S_{x_i}(\lambda) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma_{x_i}(k) e^{-i\lambda k}, \tag{5.1}$$

with $\gamma_{x_i}(k) = cov(x_{i,t}, x_{i,t-k})$, the autocovariance of x_i at lag k. The spectrum $S_{x_i}(\lambda)$ measures the variance of the cyclical component at frequency λ of the time series x_i.[4] In turn, the cross-spectrum characterizes the relationship between two time-series x_i and x_j at frequency λ

$$S_{x_i x_j}(\lambda) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma_{x_i x_j}(k) e^{-i\lambda k} = C_{x_i x_j}(\lambda) + iQ_{x_i x_j}(\lambda), \tag{5.2}$$

where $C_{x_i x_j}(\lambda)$ is the real part of the cross-spectrum and $Q_{x_i x_j}(\lambda)$ the imaginary part. Here, $\gamma_{x_i x_j}(k) = cov(x_{i,t}, x_{j,t-k})$ represents the cross-covariance between $x_{i,t}$ and $x_{j,t}$ at lag k.

Conceptually, $S_{x_i x_j}(\lambda)$ is a measure of the covariance between the cyclical components corresponding to the frequency λ of the time series $x_{i,t}$ and $x_{j,t}$. The spectra are estimated by computing first the discrete Fourier transform[5] of the time series. The squared modulus of this transform is then smoothed by a weighted moving average,[6] yielding the estimated spectrum. A detailed treatment on spectral estimation for time series can be found in Brockwell and Davis (2002), Koopmans (1995) or Warner (1998).

[4] The variance of a time series equals the total area underneath the spectrum. In other words, the spectrum shows the distribution of the total variance across the frequency band (Chatfield, 1996, p.96).

[5] The discrete Fourier transform gives the decomposition of a discrete time series x_t over its component frequencies. It is defined as $F(\lambda) = \frac{1}{\pi} \sum_{-\infty}^{\infty} x_t e^{-i\lambda t}$. The Fourier transform offers an equivalence between the frequency and the time domain.

[6] As the discrete Fourier transform is too unstable for estimation purposes, we apply a smoothed version. In our empirical application, we perform the estimation of the spectra using build-in routines of the S-Plus statistical software package. This software package applies three times the Daniell's smoother to the discrete Fourier transform (see Koopmans, 1995, for more details).

5.2.2 Dynamic Correlation

The spectral-based *dynamic correlation*, first discussed in Croux et al. (2001),[7] provides a formal measure of the correlation, or degree of co-movement, between two series x_i and x_j at each individual frequency λ, and is given by

$$\rho_{x_i x_j}(\lambda) = \frac{C_{x_i x_j}(\lambda)}{\sqrt{S_{x_i}(\lambda) S_{x_j}(\lambda)}}. \tag{5.3}$$

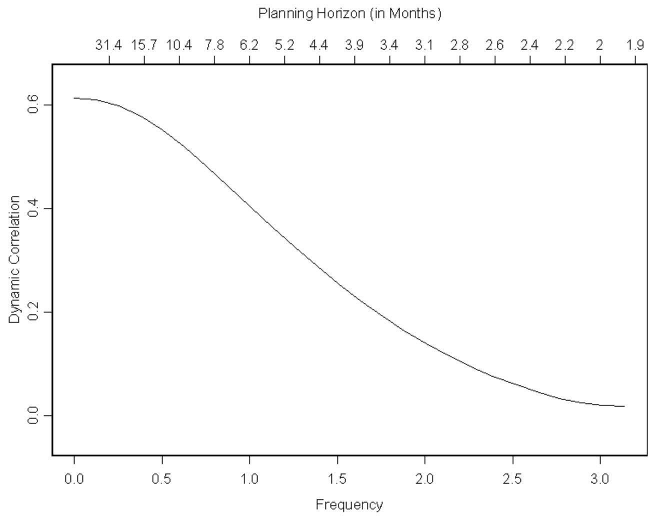
This correlation, which ranges between -1 and $+1$, is conceptually similar to the correlation between two series in the time domain. The higher its value, the more similar the fluctuations at that frequency. However, unlike the (single) static correlation in the time domain, one now obtains a correlation coefficient that can vary across different frequencies or planning horizons. Note that prior marketing studies have used the cointegration concept to describe the long-run co-movement between time series (see e.g. Franses, Kloek, & Lucas, 1999, or Srinivasan, Popkowski Leszczyc, & Bass, 2000). In so doing, one focuses on the dynamic correlation at frequency zero between the first-differentiated time series, which equals one (in absolute value) when both original series are cointegrated.[8] Our dynamic correlation concept is more comprehensive in that we look at the correlation across the entire frequency band, and not only at the zero frequency. As discussed before, the planning horizon is inversely related to the frequency. Hence, the higher (lower) the frequency, the shorter (longer) the planning horizon.

Figure 5.2 depicts graphically the estimated dynamic correlation between the aforementioned two simulated series. In line with our discussion on Figure 5.1, the lowest frequencies show the highest correlation, implying that the longer-run fluctuations in the series are strongly related,

[7] Other applications of the concept include Carlino and DeFina (2004), Partridge and Rickman (2005), Rua and Nunes (2005) and Sussmuth and Woitek (2004).

[8] The relation between cointegration and the dynamic correlation is further discussed in Croux et al. (2001).

Figure 5.2: *The dynamic correlation between the simulated series of Figure 5.1.*



i.e. show quite similar patterns. The higher frequencies correspond with a much lower correlation, implying that both series are characterized by much more idiosyncratic short-run fluctuations. Obviously, this dynamic correlation pattern is more insightful than the single static correlation coefficient of 0.293 between both simulated series.

5.2.3 Cohesion and Cross-Cohesion

From a panel of N time series, we may derive $N(N-1)/2$ possible pair-wise dynamic correlations. The higher these correlations, the more homogeneous the respective countries are, in that their customers react in a similar way to various market disturbances.

To obtain an aggregate measure of co-movement within this panel, or part of it, we can compute the *cohesion* (Croux et al., 2001) at frequency λ, denoted by $Coh(\lambda)$. For $1 \leq n_1 \leq N$ series, this cohesion is obtained as

$$Coh(\lambda) = \frac{2}{n_1(n_1-1)} \sum_{i<j}^{n_1} \rho_{x_i x_j}(\lambda). \tag{5.4}$$

Hence, the cohesion is simply the average of all possible pair-wise dynamic correlations between a given set of countries, yielding an aggregate measure of homogeneity of *correlations* over these countries. Considering our entire set of European countries ($n_1 = N$), one can thus derive an aggregate measure of European homogeneity at any given frequency. Alternatively, considering smaller subsets of countries ($n_1 < N$), one can assess the cohesion within a priori-defined country segments. In line with Tellis et al. (2001), one could, for instance, assess to what extent the Nordic, Mediterranean and Midwest segments of the European countries are more homogeneous (i.e. have a higher cohesion) than Europe as a whole, and if so, at what frequencies (planning horizons).

Apart from an aggregate measure of cohesion *within* a set of time series, one could also derive a measure of the cohesion *between* two distinct groups of time series. To that extent, one can aggregate the dynamic correlations into a *cross-cohesion* index at frequency λ,

$$Cross - Coh(\lambda) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_1+n_2} \rho_{x_i x_j}(\lambda), \tag{5.5}$$

representing the co-movement between two distinct subsets of size n_1 and n_2. In our specific setting, one could, for example, derive the cross-cohesion between the European countries and the United States, to assess whether the evolution in the European countries' CCI is in sync with the evolution in the American ICS.

The cohesion offers an aggregate measure of European homogeneity. However, there may be quite some variability between the different pair-wise dynamic correlations, which raises the question what factors drive the extent of correlation between two countries' CCI. As such, one can assess whether a larger economic, geographic and/or cultural distance significantly decreases the resulting homogeneity in the respective countries' CCI. This analysis can be implemented for specific frequencies, in which case the $N(N-1)/2$ dynamic correlations at a given frequency could be regressed against the different distance measures. Alternatively, one could average the dynamic correlations in (see Equation 5.3) over a pre-specified frequency band $\Lambda = [\lambda_1, \lambda_2[$, for $0 \leq \lambda_1 < \lambda_2 \leq \pi$, as

$$\rho_{x_i x_j}(\Lambda) = \frac{1}{(\lambda_2 - \lambda_1)} \int_{\lambda_1}^{\lambda_2} \rho_{x_i x_j}(\lambda) \mathrm{d}\lambda. \tag{5.6}$$

The heterogeneity in CCI is then analyzed by computing this average dynamic correlation over a specified frequency band $[\lambda_1, \lambda_2[$, corresponding to a time interval (planning horizon) $[T_1, T_2[$. As such, this procedure allows one to make inferences on the extent of European homogeneity across the short, medium and long run. The latter approach is conceptually similar to Deleersnyder et al. (2004), in that they also consider jointly all frequencies in a certain frequency band (in their case, all frequencies corresponding to planning horizons between two and eight years), and is less sensitive to the specific frequency one has selected. In practice, the integral in Equation 5.6 is replaced by a sum over an equally spaced grid of values of λ in the interval $\Lambda = [\lambda_1, \lambda_2[$.

To conclude, it is worth to mention that the estimation of the dynamic correlations and cohesion measures only requires the time series to be weakly stationary. No further modeling assumptions are needed.

5.3 Data

We consider the Consumer Confidence Indicator in fourteen European countries, namely Austria (AU), Belgium (BE), Denmark (DK), Finland (FI), France (FR), Germany (GE), Greece (GR), Ireland (IE), Italy (IT), Portugal (PO), Spain (SP), Sweden (SE), The Netherlands (NL), and the United Kingdom (UK). Luxembourg is not included, as no data were collected for this country before 2002. Our series span the period from November 1995, the entry date of Austria, Finland and Sweden into the European Union, until July 2005, resulting in 117 data points. The various CCI time series are depicted in Appendix A. The CCI is derived through consumer surveys collected by the European Commission and its member states in the framework of the Joint Harmonised EU Programme. Each month, over 30,000 consumers are surveyed, and the CCI is computed as the arithmetic average of the balances (in percentage points) of answers pertaining to the financial situation of the households ("How do you expect the financial position of your household to change over the next twelve months?"), the general economic situation ("How do you expect the general economic situation in this country to develop over the next twelve months?"), savings ("Over the next twelve months, how likely is it that you save any money?"), and (with inverted sign) unemployment expectations ("How do you expect the number of people unemployed in this country to change over the next twelve months?"). Respondents are asked whether they expect the variables of interest to increase, decrease, or remain stable over time. The decreases (in percentage points) are subsequently subtracted from the increases to obtain *balance* figures. A directional questionnaire is used as directional changes have been found to be easier to predict than point values (Jonung, 1986). A visual inspection of Appendix A confirms this result. These balance data are seasonally adjusted by the data provider. Details on the derivation of the CCI are provided on the website of the Directorate General Economy and Finance (DG ECFIN) of the European Commission.[9] Previous research on the CCI includes Vanden Abeele (1983), Praet and Vuchelen (1989), Batchelor and Dua (1998), and Golinelli and Parigi (2004), among others.

As differential response style (e.g. extreme ratings) can create response bias (Baumgartner and Steenkamp, 2001), the testing of construct equivalence of the CCI is highly desirable, especially in the context of cross-national research (Douglas and Craig, 2006). However, such a test is very difficult to tackle without access to the original, individual-level data. Still,

[9]http://europa.eu.int/comm/economy_finance/indicators/ businessandconsumersurveys_en.htm

we can expect the response bias to be reasonable for several reasons. First, the data are not taken in absolute levels but in balance, making the time series less sensitive to individual responses and outliers. In addition, the analysis is based on the co-movement of the national CCI, and not on the constructs themselves. As such, the problem of equivalence may be less prominent than when comparing e.g. cross-national means. Finally, some indirect evidence regarding the cross-country construct equivalence is also conveyed by the fact that similar relationships between the construct and third variables are consistently observed across multiple countries. For instance, it has been shown over a substantial number of studies, and for most European countries, that the European CCI are consistently and strongly correlated with GDP in France, Italy, Germany, and the United Kingdom (Golinelli and Parigi, 2004; Berry and Davey, 2004). Also the correlation between the CCI and household expenditures has been consistently demonstrated for the largest part of the European countries, i.e. in Belgium, Germany, France, Italy, The Netherlands, Portugal, Spain, and the United Kingdom (see Nahuis and Jansen, 2004). Finally, a consistent correlation between the European CCI and stock returns has also been found for 11 European countries, i.e. Belgium, Denmark, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain and the United Kingdom (Jansen and Nahuis, 2003).

To allow for a formal comparison with the United States, we also obtained information on the American ICS over the same time span. Following the pioneering work of Katona (1951, 1979), the ICS has been used in numerous marketing studies, such as Allenby et al. (1996), Kamakura and Gessner (1986) and Kumar, Leone, and Gaskins (1995), among others. It is not a priori clear whether the European CCI and the American ICS are conceptually equivalent, especially since the data are collected by two different institutions, i.c. the European Commission for the CCI, and the Survey Research Centre at the University of Michigan for the ICS. However, very similar types of questions are used in both surveys, even though they are not exact translations of each other.[10] In line with Croux et al. (2001), we consider four regions within the US: North-East, Midwest, South and West.

Finally, to study the cross-sectional variation in the pair-wise dynamic correlations, we introduce various distance measures. Similar to Gielens and Dekimpe (2001), the *geographic* distance between two countries i and j (GEO_{ij}) is operationalized as a dummy variable indicating whether two countries are contiguous.[11] Contiguity matrices are standard in spatial models (see e.g. Bradlow et al., 2005 or Bronnenberg, 2005). In line with Mitra and Golder (2002), the *economic* distance between two countries is based on four dimensions, i.e. the difference in the countries' economic size (reflected in their Gross Domestic Product, GDP), economic prosperity (measured through their Gross Domestic Product per Capita, $GDPC$), economic infrastructure (as reflected in the number of kilometers of railroad per square km, $RAIL$), and economic accessibility (operationalized through their population density, $DENS$).[12] The economic distance between two countries on a given dimension is defined as the absolute value of the difference between their log-transformed score on that dimension. For example, the economic-size distance is measured as

[10]For instance, the respective questions regarding the financial situation of the households are the following: (CCI) How has the financial situation of your household changed over the last 12 months? It has got a lot better, got a little better, stayed the same, got a little worse, got a lot worse, don't know? (ICS) We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago.

[11]The dummy variable takes the value zero when countries have a common border and one otherwise. Unlike the distance between capitals, this measure of contiguity does not take into account the location of the capital within the country, which can be misleading in some instances (e.g. the capital of Germany moved from Bonn to the much more remote Berlin in 1999).

[12]Other economic variables could also be used, like e.g. income inequality (Van den Bulte and Stremersch, 2004).

$|log(GDP_i) - log(GDP_j)|$. Relevant data were obtained from the World Factbook 2004.[13] To conceptualize the *cultural* distances, we use the Schwartz national-culture framework (see e.g. Schwartz, 1994; Schwartz & Ros, 1995), which has emerged as a major refinement and alternative to Hofstede's values (Steenkamp, 2001). Schwartz's framework is more recent and is based on consumer - rather than organizational - values (Steenkamp, Ter Hofstede, & Wedel, 1999) which render it more applicable to the context of our study. Cultural distance is defined in terms of the seven dimensions: conservatism ($CONS$), intellectual autonomy ($INTEL$), affective autonomy ($AFFECT$), hierarchy ($HIER$), egalitarianism ($EGAL$), harmony ($HARM$), and mastery ($MAST$). The distance on each cultural dimension is obtained as the absolute difference between two countries' score on a given dimension. Cultural data, reported in Schwartz and Ros, as well as the geographic and economic distances are available for all considered countries. As such, the regressions in the next section are implemented on 91 (= (14x13)/2) observations. All distance measures are time-invariant, as they are either intrinsically constant (geographic distance), not available as time-varying variable (cultural distance), or only collected at a higher level of temporal aggregation (economic distances) than the monthly CCI or ICS.

5.4 Results

The 14 European CCI series result in 91 possible dynamic correlations. For illustrative purposes, we first present the dynamic correlation between three key European countries: France, Germany and the United Kingdom. Next, we derive an aggregate measure for the degree of homogeneity across the different member states through the cohesion index , and compare this measure with (i) the cohesion in ICS across the four US regions, and (ii) the cross-cohesion between the US and the European Union. We subse-

[13] Available on the website of the CIA http://www.cia.gov/cia/publications/factbook/.

quently assess whether there are certain clusters of countries which, among themselves, are relatively more homogeneous than the Union as a whole. Finally, we assess whether the observed variability between the pair-wise dynamic correlations is driven by the geographic, economic and/or cultural distance(s) between the respective countries, and how this relative importance varies across different planning horizons.

5.4.1 Pair-Wise Dynamic Correlations

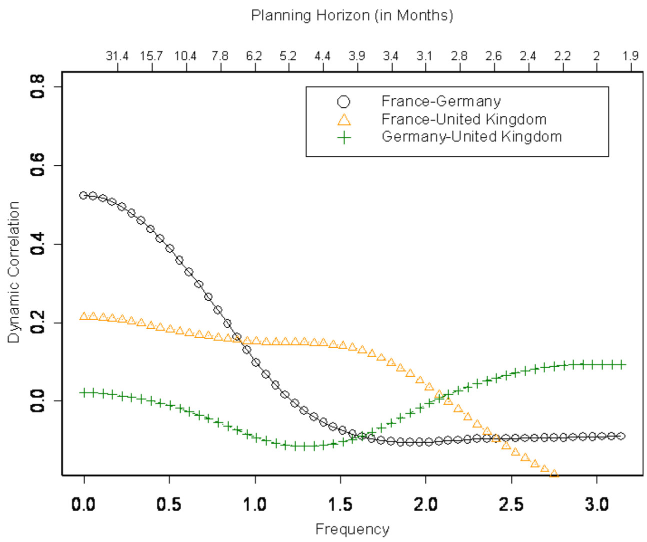
Rather than presenting all 91 dynamic correlations (which are available from the authors upon request), we focus on the dynamic correlations between the CCI of three key countries: France, Germany and the United Kingdom. France and Germany are often seen as two key forces (both economically and politically) of the European Unification (The Economist, 2003). The United Kingdom, in contrast, while also being an important player, has been argued to have a rather distinct position, not only geographically, but also in terms of economic integration and culture (Nothcott, 1995).

In line with Jansen and Nahuis (2003), preliminary unit-root tests found the different CCI series to be integrated of order one.[14] The dynamic correlations were therefore computed on the first differences. For notational simplicity, we still refer to these first-differenced series as CCIs. The corresponding dynamic correlations are presented in Figure 5.3. On the bottom horizontal axis, we depict the frequency in radians, while the top axis presents the corresponding planning horizon (in months). As indicated before, the higher the frequency, the lower the planning horizon. In all instances, the short-run dynamic correlation (corresponding with the higher frequencies) is close to zero. This suggests that many of the disturbances that drive the high-frequency (monthly, bimonthly, etc) fluc-

[14] Note, however, that pair-wise Johansen cointegration tests indicate that the series are not cointegrated. Results are available from the authors upon request.

tuations in consumer confidence are country specific and not correlated across the respective countries. This short-run heterogeneity supports the idea of multi-domestic strategies. However, especially in the case of France and Germany, this may be an overly myopic view, in that the dynamic correlation increases considerably as the planning horizon is extended beyond six months. Market shocks that drive the longer-evolution in consumers' confidence therefore have a similar impact in both countries, which supports a more integrated approach across these two countries. The dynamic correlations with the United Kingdom, in contrast, remain considerably smaller at all frequencies. These findings, based on consumer perceptions, are in line with earlier research by Lemmens, Croux, and Dekimpe (2005). In their pan-European study on the predictive content of managers' production expectations, they found significant cross-border effects between France and Germany, while the UK occupied a fairly isolated position.

Figure 5.3: *Dynamic correlation for France, Germany and the United Kingdom.*



While one should be careful when generalizing from a limited number of cases, the above discussion already suggests that there is little homogeneity in the short-run fluctuations in consumers' confidence. In terms of the longer-run movements, in contrast, there seems to be more variability across country pairs, and a potential to identify relatively homogeneous subsets. Finally, the observed differences seem to be related to the relative "closeness" of the different countries. Next, we investigate more formally these preliminary patterns.

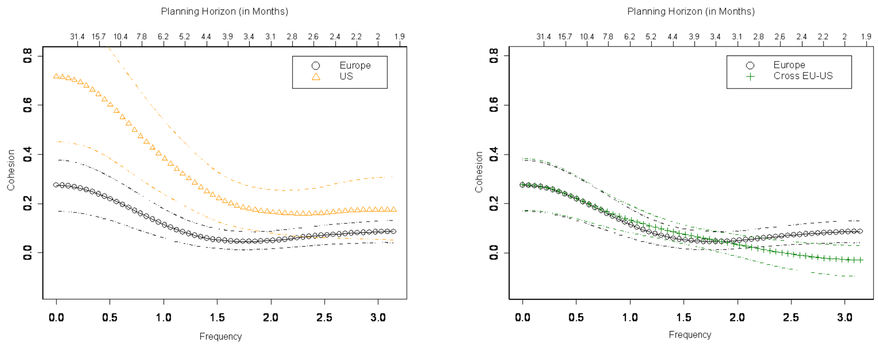
5.4.2 The European Cohesion in Consumer Confidence

The first set of observations is confirmed by computing the European cohesion measure, which aggregates all 91 pair-wise correlations. Figure 5.4 (left panel) presents the estimated cohesion, along with 90% bootstrap confidence bands.[15] As indicated in the figure, the European cohesion is very low at the high frequencies, suggesting very little pan-European homogeneity in the short-run fluctuations in consumer confidence across the different member states. This implies that either country-specific shocks (local unemployment figures, the outcome of local elections, etc) drive these short-run fluctuations, or that different countries have different short-run reactions to common shocks (s.a. news issued by the European Central Bank, world events, etc). Illustrating the former case, the closure of Renault's Belgian factory, announced in February 1997 (The Economist, 1997), caused a sharp fall in the Belgian CCI of 7 points, while most other countries were unaffected. The common shock of September 11, 2001 in turn, affected the confidence in all member states considerably, but some countries (e.g. the British and Irish CCIs lost 7 points over the month) to

[15] Confidence bounds around the estimated (cross-)cohesion measures are computed using non-parametric block-bootstrap, as in Croux et al. (2001). An overview of bootstrap methods for time series can be found in Davison and Hinkley (2003). In this application, the block-bootstrap is implemented with blocks of minimum length 12, and standard errors are obtained from 1000 bootstrap replications of the cohesion measure.

a much larger extent than others (e.g. the Nordic countries lost less than 2 points).[16]

Figure 5.4: *Cohesion within Europe and within the United States (left panel) and cross-cohesion between Europe and US and cohesion within Europe (right panel).*



In line with the patterns observed for France and Germany, we further see that the cohesion increases somewhat as the planning horizon is extended, indicating a more homogeneous evolution once the dust has settled. To put the European cohesion levels in perspective, we compute as benchmark the cohesion in the ICS across the four US regions (see Figure 5.4, left panel).[17] A priori, we expect the latter cohesion to be considerably higher, if only because the United States have a much longer history of unification while also sharing a common language and currency, and having a single foreign policy and army. Across the entire range of frequencies, the US-based cohesion exceeds its European counterpart. This difference is statistically significant for planning horizons beyond 3.4 months. Interestingly, at the higher frequencies, we see that also within the United States, there remains considerable heterogeneity in the behavior of the ICS. This

[16]More details can be found in the “Employment in Europe 2001, Autumn Update” report of the European Commission, DG Employment and Social Affairs.

[17]The absence of ICS data at the state level precludes the derivation of a (cross-) cohesion measure for the 52 states of America.

finding is in line with the work of Wells and Reynolds (1979) and Hawkins, Roupe and Coney (1981) who found significant geographical variation in consumer values, attitudes and consumption across different regions of the United States, and of Mittal, Kamakura and Govind (2004), who found such differences in consumers' satisfaction with car dealers. However, because of the cross-sectional nature of their data, the increasing homogeneity over longer time horizons could not be inferred from these earlier studies.

When looking at the cross-cohesion between Europe and the different US regions (see Figure 5.4, right panel along with 90% bootstrap confidence bands), we find a comparable pattern, with higher correlations at lower frequencies. As the planning horizon is extended, the European CCI and the American ICS increasingly react in similar ways. While this may not seem too surprising given the United States' economic and political power in today's global marketplace (Julius, 2005), it is interesting to note that the cohesion within Europe does not significantly exceed the cross-cohesion level. That is, average correlations between pairs of European countries and between European countries and US regions turn out to be of a comparable magnitude, in particular at the medium and longer run. As a potential reason, Europe and the United States are each other's main trading partners, both accounting for around one fifth of each other's bilateral trade, a matter of 1 billion a day.[18] Hence, recent political claims on Europe's distinct (relative to the US) identity are not yet fully reflected in its consumers' perceptions.

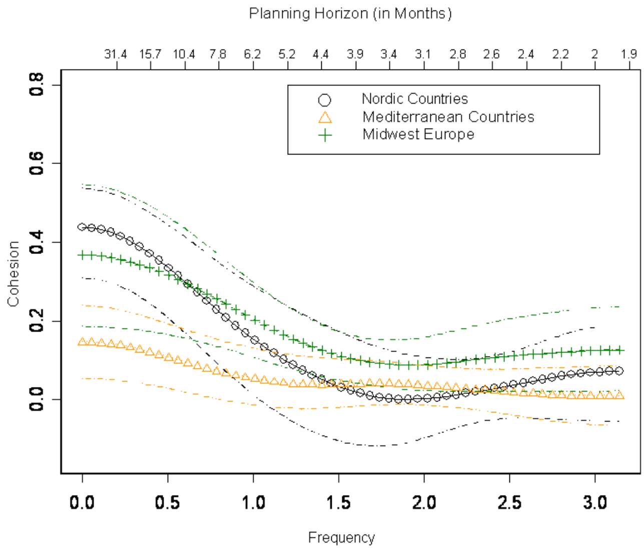
5.4.3 European Segments

As the overall cohesion across all 14 countries is fairly small, even at the lower frequencies, the question emerges whether this picture changes when considering smaller subsets of countries. Indeed, a few discrepant countries

[18] See the European Commission's official website http://europa.eu.int/comm/trade/issues/bilateral/countries/usa/index_en.htm.

may well drive the overall homogeneity estimate down. Looking at Figures 5.2 and 5.3, it is obvious that the European-based cohesion is considerably lower than the dynamic correlations reported between France and Germany. While one could adopt several a priori segmentation schemes, we follow the typology adopted in Tellis et al. (2003), and identify the following three segments: (i) the Nordic (DK, FI, and SE), (ii) Mediterranean (FR, GR, IT, PO, and SP), and (iii) Midwest countries (AU, BE, GE, IE, NL, and UK).

Figure 5.5: *The cohesion within predefined market segments.*



As indicated in Figure 5.5, especially the Nordic and Midwest countries are characterized by a considerably higher homogeneity at the lower frequencies. The cohesion at longer planning horizons within the Nordic and the Midwest segments approaches the values obtained within the United States (that is, their confidence bands overlap for small frequencies). The emergence of a homogeneous Nordic segment confirms previous findings of Kumar et al. (1998), Helsen, Jedidi, and DeSarbo (1993) and Tellis et al. (2003). Much less homogeneity is observed among the Mediterranean countries, irrespective of the time horizon considered. These findings are in line with Bijmolt et al. (2004) who, in their study on financial-product ownership, identified relatively homogeneous segments among respectively, the Nordic and Midwest countries, while most Mediterranean countries formed single-country segments. Again, very little cohesion is observed at short planning horizons, irrespective of the country segment.

5.4.4 Does Distance Still Matter?

The examples in Figure 5.3 (for France, Germany and the UK) suggested that there may be quite some variability in dynamic correlation both across different country pairs, and across different time horizons. To more formally assess this variability, we regress the pair-wise correlations across various indicators of economic, geographic and cultural distance, for three different planning horizons, i.e. the short, medium and longer run.

In the marketing literature, no unique definition exists as to what constitutes the short, medium and long run (see in this respect the very different operationalizations advocated in Dekimpe & Hanssens, 1999, and Mela, Gupta, & Lehmann, 1997). As it has been found that consumers' attitudes change quickly (Leone & Kamakura, 1983), causing them to sometimes use very short (even monthly) planning horizons (Thaler, 1985), we define our short-run planning horizon as those fluctuations with a periodicity inferior to four months. This corresponds to a frequency band $\Lambda_1 = [\frac{\pi}{2}, \pi[$. The medium term is assumed to correspond to a planning horizon of four to twelve months, with frequency band $\Lambda_2 = [\frac{\pi}{6}, \frac{\pi}{2}[$, while the longer-term fluctuations are assumed to correspond with cycles of twelve months to two years, i.e. frequency band $\Lambda_3 = [\frac{\pi}{12}, \frac{\pi}{6}[$. We do not take fluctuations of lower frequency into account, to ensure a sufficient number of cycles for reliable analysis.[19] As indicated previously, we integrate the dynamic cor-

[19]We observe nearly five cycles of two years in our sample. We also checked for the

relations over the different frequencies in a given frequency band to arrive at a single (average) estimate for the dynamic correlation in that band.

Three regression models are subsequently estimated, with as dependent variable the dynamic correlation in, respectively, the short, medium and long-run frequency band, and as explanatory variables the various indicators for geographic (GEO), economic (GDP, $GDPC$, DEN, $RAIL$) and cultural ($CONS$, $HIER$, $AFFEC$, $INTEL$, $MAST$, $EGAL$, $HARM$) distance. Fisher z-transforms are applied to the dynamic correlations, since the latter are typically not normally distributed. Single-equation estimation techniques are used. A system's approach would not result in more efficient parameter estimates, as all equations contain the same set of explanatory variables. Preliminary White tests (available upon request) do not reveal significant heteroskedasticity in any of the regressions. As each observation in the regressions corresponds to a pair of countries, possible correlation among the error terms can be modeled by introducing random country effects, as in Sethuraman, Srinivasan, and Kim (1999). The latter, however, turned out not to be important.[20] Hence, we preferred to stick to the OLS estimator. Finally, as there may be multicollinearity between the different indicators of economic (cultural) distance, we focus on the more robust joint p-values.[21] These are reported in Table 5.1. The individual

robustness of our results when modifying the spans of the short-, medium- and long-run planning horizons. Results turn out to be highly robust, and conclusions to be maintained for variations in the definition of the intervals.

[20]Farley and Lehmann (1986) note in this respect that the bias due to non-independence may not be serious if the percentage of non-zero correlations between pairs of error terms is relatively small. In our application, this ratio is about 15%. When adopting a GLS approach to account for the aforementioned dependencies, qualitatively similar conclusions were indeed obtained (detailed results available upon request).

[21]There are 91 observations for 12 explanatory variables, some of them highly correlated. Given the potential multicollinearity, individual parameter estimates are unstable and difficult to interpret. By testing for the joint significance of the three groups of covariates, much more stable joint p-values are obtained, identifying the total impact of the groups of explanatory variables (Verbeek, 2000, p.40). Alternatively, a composite

coefficient estimates can be found in Appendix B.

Table 5.1: *OLS-estimated p-values, F-statistics and R^2 measures for different time horizons.*

	Static correlation	Short run	Middle run	Long run
Joint test p-values				
Geographic distance	0.307	0.749	0.237	0.078
Economic distances	0.130	0.438	0.081	0.009
Cultural distances	0.246	0.300	0.137	0.047
Overall F-statistic, p-value	0.064	0.256	0.019	0.001
Adjusted R^2	0.096	0.034	0.138	0.242

Remember that in terms of the short-run correlations, very small values were obtained for each of the three country pairs of Figure 5.3. This pattern was also found in the larger set of correlations. Not surprisingly, the short-run regression results in a very low adjusted[22] R^2 (0.034), and an insignificant overall F-statistics ($p = 0.256$). This could partially be explained by the fact that measurement error in the construct is picked up as short-run fluctuation, and these measurement errors are not related to the distances between countries. Irrespective of the geographic, economic or cultural distance, the high-frequency fluctuations in two countries' CCI do not show much correlation.

As one moves to the lower-frequency movements in CCIs, the explanatory power of the cross-sectional regressions increases. In the medium run, the adjusted R^2 increases to 0.138, and becomes 0.242 in the long run.

index of cultural (economic) distance can be constructed (as in Kogut & Singh, 1988) in order to reduce the number of variables. However, we believe that joint p-values are more powerful, since replacing explicative variables by a composite index leads to information loss.

[22]Given the high number of explanatory variables relative to the number of observations, the use of adjusted R^2 is more appropriate.

Also the corresponding F-statistics become highly significant ($p = 0.019$ and 0.001, respectively). In the medium run, the economic distance becomes significant ($p < 0.10$), while in the long run, all three distance components become significant. The correlation in longer-run CCI movements decreases as the geographic distance increases, as the economic distance becomes larger,[23] and as countries become more culturally different. No such insights could have been obtained from the traditional static correlations, as this resulted in a poorly fitting (adjusted $\mathrm{R}^2 = 0.096$), with a weak overall significance (p-value of the overall F-statistics $= 0.064$), and none of the three distances being significant.

5.5 Conclusions

The ongoing unification which takes place on the European political scene, along with recent advances in consumer mobility and communication technology, raises the question whether the different member states of the European Union can be treated as a single market to take full advantage of pan-European marketing strategies. However, *distance remains an important determinant* of (dis)similarities in European consumers' confidence. Recent claims on the "death of distance" (The Economist, 1995) are therefore premature.

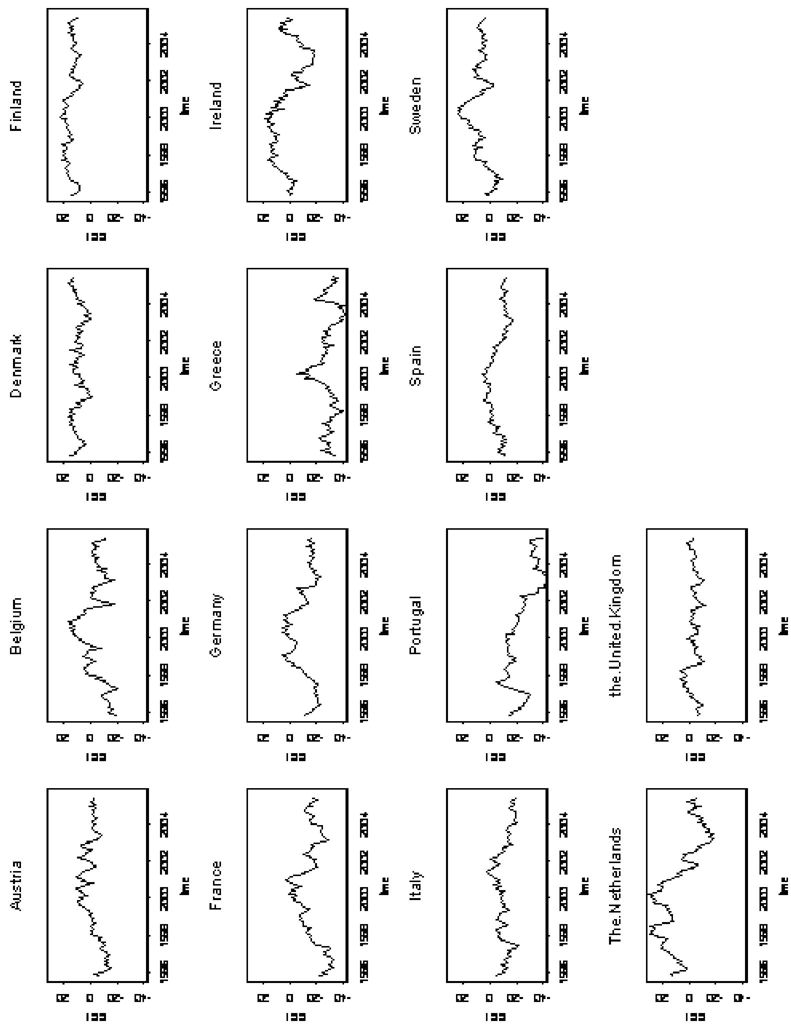
Our analyses clearly indicate that the European Union does not yet constitute a single, homogeneous, market. Not only are the short-run (high-frequency) movements in consumers' confidence driven by country-specific shocks and/or differing reactions to common shocks, but also the homogeneity in their longer-run reactions decreases significantly as the

[23]Indeed, the sum of the coefficients of the economic distances is negative (see Appendix B). This means that, if every type of economic distance increases with one unit, we get a total significant negative effect on the dynamic correlations. Note that the economic distances are all measured on the same ratio-scale. The same applies for the seven cultural distances.

distance between the different European countries increases. As such, in terms of short-term tactical marketing decision making, country-specific strategies may still be called for. For more strategic decisions that have longer-run implications, there is more cross-country homogeneity to exploit, but the continued significance of geographic, economic and cultural distances suggests more potential for pan-regional strategies than for a single pan-European strategy.

This study has several limitations offering areas for future research. The most important limitation it that the CCI is the only construct we use to measure differences between consumers. It may be fruitful to study the generalizability of our findings for other constructs as segmentation bases. Second, the construct and measure equivalence (e.g. Steenkamp & Ter Hofstede 2002) of the CCI is not established yet. While this issue was out of the scope of the present study, and many other aforementioned references also using this construct, it offers an interesting area for future cross-national research. We hope that this study will help spark further research on the unity of the European market.

Appendix A: *The evolution of the CCIs in Europe.*



Appendix B: *OLS-estimated regression coefficients (and their standard errors) of the drivers of variability among the dynamic correlations for different time horizons.*

	Static correlation	Short run	Middle run	Long run
Geographic distance				
GEO	-0.033	0.014	-0.051	-0.099*
	(0.032)	(0.042)	(0.043)	(0.056)
Economic distances				
GDP	-0.003	0.017	-0.027	-0.052**
	(0.013)	(0.016)	(0.017)	(0.022)
GDPC	-0.051	-0.074	-0.096	-0.084
	(0.092)	(0.121)	(0.122)	(0.159)
DEN	0.037**	0.022	0.042*	0.077**
	(0.018)	(0.023)	(0.023)	(0.030)
RAIL	-0.034	-0.033	-0.014	-0.034
	(0.027)	(0.035)	(0.035)	(0.046)
Joint test p-value	0.130	0.438	0.081*	0.009***
Cultural distances				
CONS	-0.026	-0.047	0.099	0.169
	(0.121)	(0.159)	(0.160)	(0.209)
HIER	0.150*	0.127	0.264***	0.227*
	(0.075)	(0.099)	(0.100)	(0.130)
AFFEC	0.035	0.107	-0.086	-0.189*
	(0.055)	(0.073)	(0.073)	(0.096)
INTEL	-0.008	-0.021	0.000	0.107
	(0.080)	(0.105)	(0.105)	(0.137)
MAST	-0.096	-0.143	-0.041	-0.198
	(0.115)	(0.151)	(0.152)	(0.198)
EGAL	-0.116	-0.043	-0.146	-0.227
	(0.108)	(0.142)	(0.143)	(0.186)
HARM	-0.126**	-0.076	-0.121	-0.135
	(0.060)	(0.078)	(0.079)	(0.103)
Joint test p-value	0.246	0.300	0.137	0.047**
Intercept	0.127***	0.063	0.150**	0.331***
	(0.046)	(0.061)	(0.061)	(0.080)
$n = 91$				
Overall F-statistic, p-value	0.064*	0.256	0.019**	0.001***
Adjusted R^2	0.096	0.034	0.138	0.242

* $p < 0.100$; ** $p < 0.050$; *** $p < 0.010$.

References

[1] Adams, F.G. (1965). Prediction with consumer attitudes: The time series/cross-section paradox. *Review of Economics and Statistics*, 47(4), 367-378.

[2] Allenby, G.M., Jen, L., & Leone, R.P. (1996). Economic trends and being trendy: The influence of consumer confidence on retail fashion sales. *Journal of Business and Economic Statistics*, 14(1), 103-111.

[3] Askegaard, S., & Madsen, T.K. (1998). The local and the global: exploring traits of homogeneity and heterogeneity in European food cultures. *International Business Review*, 7(6), 549-568.

[4] Barksdale, H.C., Hilliard, J.E., & Ahlund, M.C. (1975). A cross-spectral analysis of beef prices. *The Journal of Agricultural Economics*, May, 309-315.

[5] Barksdale, H.C., Hilliard, J.E., & Guffey, H.J. (1974). Spectral analysis of advertising-sales interaction: An illustration. *Journal of Advertising*, 3(4), 26-33.

[6] Batchelor, R., & Dua, P. (1998). Improving macro-economic forecasts: The role of consumer confidence. *International Journal of Forecasting*, 14(1), 71-81.

[7] Baumgartner, H. & Steenkamp, J.-B.E.M. (2001). Response styles in marketing research: A cross-national investigation, *Journal of Marketing Research*, 38(May), 143-156.

[8] Baxter, M. (1994). Real exchange rates and real interest differentials. Have we missed the business-cycle relationship? *Journal of Monetary Economics*, 33, 5-37.

[9] Berry, S., & Davey, M. (2004). How should we think about consumer confidence? *Bank of England Quarterly Bulletin*, Autumn, 282-290.

[10] Bijmolt, T.H.A., Paas, L.J., & Vermunt, J. (2004). Country and consumer segmentation: Multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing*, 21(4), 323-340.

[11] Boote, A.S. (1983). Psychographic segmentation in Europe. *Journal of Advertising Research*, 22(December), 19-25.

[12] Bradlow, E.T., Bronnenberg, B.J., Russell, G.J., Arora, N., Bell, D.R., Duvvuri, S.D., Ter Hofstede, F., Sismeiro, C., Thomadsen, R., & Yang, S. (2005). Spatial models in marketing. *Marketing Letters*, 16(3-4), 267-278.

[13] Brockwell, P.J., & Davis, R.A. (2002). *Introduction to time series and forecasting* (2nd ed.), Chapter 4, New York: Springer Texts in Statistics.

[14] Bronnenberg, B.J., Mela, C., & Boulding, W. (2004). The Periodicity of competitor pricing. *Working paper series at the Andersen School of Management*, UCLA.

[15] Bronnenberg, B.J. (2005). Spatial models in marketing research and practice. *Applied Stochastic Models in Business and Industry*, 21(4-5), 335-343.

[16] Burch, S.W., & Gordon, S. (1984). The Michigan surveys and the demand for consumers durables. *Business Economics*, 19(1), 40-44.

[17] Carlino, G.A., & DeFina, R.H. (2004). How strong is co-movement in employment over the business cycle? Evidence from state/sector data. *Journal of Urban Economics*, 55(2), 298-315.

[18] Chatfield, C. (1996). *The analysis of time series: An introduction* (5th Ed.). London: Chapman & Hall.

[19] Chauvet, M., & Guo, J.T. (2003). Sunspots, animal spirits and economic fluctuations. *Macroeconomic Dynamics*, 7(1), 140-169.

[20] Croux C., Forni M., & Reichlin, L. (2001). A measure of comovement for economic variables: Theory and empirics. *The Review of Economics and Statistics*, 83(2), 232-241.

[21] Davison, A.C., & Hinkley, D.V. (2003). *Bootstrap methods and their application.* Cambridge: Cambridge University Press.

[22] Dekimpe, M.G., & Hanssens, D.M. (1999). Sustained spending and persistent response: A new look at long-term marketing profitability. *Journal of Marketing Research*, 36(4), 397-412.

[23] Dekimpe, M.G., & Hanssens, D.M. (2004). Persistence modeling for assessing marketing strategy performance. In C. Moorman & D.R.Lehmann (Eds.), *Assessing marketing strategy performance*, Cambridge: MSI Institute.

[24] Deleersnyder B., Dekimpe, M.G., & Leeflang, P.S.H. (2004). Advertising and the macro-economy: Have we missed the business-cycle relationship? *Paper presented at the 33th EMAC Conference*, Murcia.

[25] Deleersnyder, B., Dekimpe, M.G., Sarvary, M., & Parker, P.M. (2004). Weathering tight economic times: The sales evolution of consumer durables over the business cycle. *Quantitative Marketing and Economics*, 2(4), 347-383.

[26] De Mooij, M., & Hofstede, G. (2002). Convergence and divergences in consumer behavior: Implications for international retailing? *Journal of Retailing*, 78(1), 61-69.

[27] Douglas, S.P., & Craig, C.S. (2006). On improving the conceptual foundations of international marketing research. *Journal of International Marketing*, 14(1), 1-22.

[28] Farley, J.U., & Lehmann, D.R. (1986). *Meta-analysis in marketing: Generalization of response models*. Lexington, MA: Lexington Press.

[29] Franses, P.H., Kloek, T., & Lucas, A. (1999). Outlier robust analysis of long-run marketing effects for weekly scanning data. *Journal of Econometrics*, 89(1), 293-315.

[30] Friend, I., & Adams, F.G. (1964). The predictive ability of consumer attitudes, stock prices, and non-attitudinal variables. *Journal of the American Statistical Association*, 59(308), 987-1005.

[31] Gielens, K., & Dekimpe, M.G. (2001). Do international entry decisions of retail chains matter in the long run? *International Journal of Research in Marketing*, 18(3), 235-259.

[32] Gielens, K., & Steenkamp, J.-B.E.M. (2004). What drives new product success? An investigation across products and countries. *MSI report* 04-002, 25-49.

[33] Golinelli, R., & Parigi, G. (2004). Consumer sentiment and economic activity: A cross-country comparison. *Journal of Business Cycle Measurement and Analysis*, 1(2), 147-170.

[34] Hawkins, D.I., Roupe, D, & Coney, K.A. (1981). The influence of geographical subcultures in the United States. *Advances in Consumer Research*, 8(1), 713-7.

[35] Helsen, K., Jedidi, K., & DeSarbo, W.S. (1993). A new approach to country segmentation utilizing multinational diffusion patterns. *Journal of Marketing*, 57(4), 60-71.

[36] Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* Beverly Hills: Sage Publications.

[37] Holt, D.B., Quelch, J.A., & Taylor, E.L. (2004). How global brands compete. *Harvard Business Review*, 82(9), 68-75.

[38] Jansen, W.J., & Nahuis, N.J. (2003). The stock market and consumer confidence: European evidence. *Economics Letters*, 79(1), 89-98.

[39] Jonung, L. (1986). Uncertainty about inflationary perceptions and expectations. *Journal of Economic Psychology*, 7(3), 315- 325.

[40] Julius, D. (2005). US economic power. *Harvard International Review*, 26(4), 14-18.

[41] Kamakura, W.A., & Gessner, G. (1986). Consumer sentiment and buying intentions revisited: A comparison of predictive usefulness. *Journal of Economic Psychology*, 7(2), 197-220.

[42] Kamakura, W.A., Novak, T.P., Steenkamp, J.-B.E.M., & Verhallen, T.M.M. (1994). Identification de segments de valeurs pan-europens par un modle logit sur les rangs avec regroupements successifs, *Recherche et Applications en Marketing*, 8(4), 30-55.

[43] Kumar, V., Stam, A., & Joachimsthaler, E.A. (1994). An interactive multi-criteria approach to identifying potential foreign markets. *Journal of International Marketing*, 2(1), 29-52.

[44] Katona, G. (1951), *Psychological analysis of the economic behaviour.* New York: McGraw-Hill.

[45] Katona, G. (1979). Towards a macropsychology. *American Psychologist*, 34, 118-126.

[46] Koopmans, L.H. (1995). *The spectral analysis of time series.* Probability and Mathematical Statistics (Vol. 22), San Diego, California: Academic Press.

[47] Kraus, P.A. (2003). Cultural pluralism and European policy-building. *Journal of Common Market Studies*, 41(4), 665-686.

[48] Kumar V., Leone, R.P., & Gaskins, J.N. (1995). Aggregate and disaggregate sector forecasting using consumer confidence measures. *International Journal of Forecasting*, 11(3), 361-377.

[49] Kumar, V., Ganesh, J., & Echambadi, R. (1998). Cross-national diffusion research: What do we know and how certain are we? *Journal of Product Innovation Management*, 15(3), 255-268.

[50] Lemmens, A., Croux, C., & Dekimpe, M.G. (2005). On the predictive content of production surveys: A pan-European study. *International Journal of Forecasting*, 21(2), 363-375.

[51] Leone, R.P., & Kamakura, W.A. (1983). Usefulness of indices of consumer sentiment in predicting expenditures. In R.P. Bagozzi, & A.M. Tybout (Eds.), *Advances in consumer research* (pp. 596-599), Ann Arbor, MI: Association for Consumer Research.

[52] Mahajan, V., & Muller, E. (1994). Innovation diffusion in a borderless global market: Will the 1992 unification of the European Community accelerate diffusion of new ideas, products, and technologies? *Technological forecasting and Social Change*, 45(3), 221-235.

[53] Mc Donald, F., & Dearden, S. (2005). *European economic integration* (4th Ed.). London: Prentice Hall.

[54] Mela, C.F., Gupta, S., & Lehmann, D.R. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research*, 34(2), 248-261.

[55] Mitra, D., & Golder, P.N. (2002). Whose culture matters? Near-market knowledge and its impact on foreign market entry timing. *Journal of Marketing Research*, 39(3), 350-365.

[56] Mittal, V., Kamakura, W.A., & Govind, R. (2004). Geographic patterns in customer satisfaction: An empirical investigation. *Journal of Marketing*, 68(3), 48-62.

[57] Mueller, E. (1963). Ten years of consumer attitude surveys: Their forecasting record. *Journal of the American Statistical Association*, 58(304), 899-917.

[58] Nahuis, N.J., & Jansen, W.J. (2004). Which survey indicators are useful for monitoring consumption? Evidence from European countries. *Journal of Forecasting*, 23, 89-98

[59] Nijs, V.R., Dekimpe, M.G., Steenkamp, J.-B.E.M., & Hanssens, D.M. (2001). The category-demand effects of price promotions. *Marketing Science*, 20(1), 1-22.

[60] Northcott, J. (1995). *The future of Britain and Europe.* London Policy Studies Institute Research Report.

[61] Özsomer, A., & Simonin, B.L. (2004). Marketing program standardization: A cross-country exploration. *International Journal of Research in Marketing*, 21(4), 397-419.

[62] Parsons, L.J., & Henry, W.A. (1972). Testing equivalence of observed and generated time series data by spectral methods. *Journal of Marketing Research*, 9(4), 391-395.

[63] Partridge, M.D., & Rickman, D.S. (2005). Regional cyclical asymmetries in an optimal currency area: An analysis using US state data. *Oxford Economic Papers*, 57(3), 373-397.

[64] Pauwels, K., Currim, I., Dekimpe, M.G., Ghysels, E., Hanssens, D.M., Mizik, N., & Naik, P. (2005). Modeling marketing dynamics by time series econometrics. *Marketing Letters*, 15(4), 167-183.

[65] Pauwels, K., Hanssens, D.M., & Siddarth, S. (2002). The long-term effects of price promotions on category incidence, brand choice, and purchase quantity. *Journal of Marketing Research*, 39(4), 421-439.

[66] Praet, P., & Vuchelen, J. (1989). The contribution of consumer confidence indexes in forecasting the effects of oil prices on private consumption. *International Journal of Forecasting*, 5(3), 393-397.

[67] Priestley, M.B. (1981). *Spectral analysis and time series.* New York: Academic Press.

[68] Putsis, W.P., Balasubramanian, J.S., Kaplan, E.H., & Sen, S.K. (1997), Mixing Behavior in Cross-Country Diffusion. *Marketing Science*, 16 (4), 354-369.

[69] Rosenberger, S. (2004). The other side of the coin. *Harvard International Review*, 26(1), 22-25.

[70] Rua A., & Nunes, L.C. (2005). Coincident and leading indicators for the euro area: A frequency band approach. *International Journal of Forecasting*, 21(3), 503-523.

[71] Schwartz, S.H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, H.C. Triandis, C. Kgitibasi, S.-C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (pp.85-119), Thousand Oaks, CA: Sage.

[72] Schwartz, S.H., & Ros, M. (1995). Values in the West: A theoretical and empirical challenge to the individualism-collectivism cultural dimension. *World Psychology*, 1(2), 91-122.

[73] Sethuraman, R., Srinivasan, V., & Kim, D. (1999). Asymmetric and neighborhood cross-price effects: Some empirical generalizations. *Marketing Science*, 18(1), 23-41.

[74] Srinivasan, S., Popkowski Leszczyc P.T.L., & Bass, F.M. (2000). Market share response and competitive interaction: The impact of temporary, evolving and structural changes in prices. *International Journal of Research in Marketing*, 17(4), 281-305.

[75] Steenkamp, J.-B.E.M. (2001). The role of national culture in international marketing research. *International Marketing Review*, 18(1), 30-44.

[76] Steenkamp, J.-B.E.M., & Ter Hofstede, F. (2002). International market segmentation: Issues and perspectives. *International Journal of Research in Marketing*, 19(3), 185-213.

[77] Steenkamp, J.-B.E.M., Ter Hofstede, F., & Wedel, M. (1999). A cross-national investigation into the individual and national cultural antecedents of consumer innovativeness. *Journal of Marketing*, 63 (2), 55-69.

[78] Stremersch, S., & Tellis, G.J. (2004). Understanding and managing international growth of new products. *International Journal of Research in Marketing*, 21(4), 421-438.

[79] Sussmuth, B., & Woitek, U. (2004). Business cycles and comovement in Mediterranean economies - A national and area wide perspective. *Emerging Markets Finance and Trade*, 40(6), 7-27.

[80] Tellis, G.J., Stremersch, S., & Yin, E. (2003). The international takeoff of new products: The role of Economics, culture, and country innovativeness. *Marketing Science*, 22(2), 188-208.

[81] Ter Hofstede, F., Steenkamp, J.-B.E.M., & Wedel, M. (1999). International market segmentation based on consumer-product relations. *Journal of Marketing Research*, 36(1), 1-17.

[82] Ter Hofstede, F., Wedel, M. , & Steenkamp, J.-B.E.M. (2002). Identifying spatial segments in international markets. *Marketing Science*, 21(2), 160-177.

[83] Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3), 199-214.

[84] The Economist (1995, September 30). The death of distance. Telecommunications survey. 336, 5-6.

[85] The Economist (1997, March 8). Europe's great car war. 342, 69-70.

[86] The Economist (1999, October 2). Converging by diverging. 353, 84-87.

[87] The Economist (2003, October 25). The Franco-German monster. 369, 50.

[88] The European Voice (2001, July 26). Car pricing study labelled 'unfair' by manufacturers. 7.

[89] Throop, A.W. (1992). Consumer sentiment: Its causes and effects. *Federal Reserve Bank of San Francisco, Economic Review*, 1, 35-59.

[90] Vanden Abeele, P. (1983). The index of consumer sentiment: Predictiability and predictive power in the EEC. *Journal of Economic Psychology*, 3(1), 1-17.

[91] Van den Bulte, C. & Stremersch, S. (2004). Social contagion and income inequality in new product diffusion: a meta-analytic test. *Marketing Science*, 23(4), 530-544.

[92] Verbeek, M. (2000). *A guide to modern econometrics.* Chichester: John Wiley & Sons.

[93] Warner, R.M. (1998). *Spectral Analysis of Time-Series Data.* New York: The Guilford Press.

[94] Wedel, M., & Kamakura, W.A. (1998). *Marketing segmentation: Conceptual and methodological foundations.* Boston: Kluwer Academic Publishers.

[95] Wells, W.D., & Reynolds, F.D. (1979). Psychological geography. In J.N. Sheth, (Ed.), *Research in marketing*, 2 (pp. 345-357), JAI Press.

[96] Yip, G.S. (1995). *Total global strategy.* New Jersey: Prentice-Hall.

List of Tables

List of Figures

Doctoral dissertations from the Faculty of Economics and Applied Economics

From August 1, 1971.

1. GEPTS Stefaan (1971)
   Stability and efficiency of resource allocation processes in discrete commodity spaces. Leuven, K. U. Leuven, 1971. 86 pp.
2. PEETERS Theo (1971)
   Determinanten van de internationale handel in fabrikaten. Leuven, Acco, 1971. 290 pp.
3. VAN LOOY Wim (1971)
   Personeelsopleiding: een onderzoek naar investeringsaspekten van opleiding. Hasselt, Vereniging voor wetenschappelijk onderzoek in Limburg, 1971. VII, 238 pp.
4. THARAKAN Mathew (1972)
   Indian exports to the European community: problems and prospects. Leuven, Faculty of Economics and Applied Economics, 1972. X, 343 pp.
5. HERROELEN Willy (1972)
   Heuristische programmatie: methodologische benadering en praktische toepassing op complexe combinatorische problemen. Leuven, Aurelia scientifica, 1972. X, 367 pp.
6. VANDENBULCKE Jacques (1973)
   De studie en de evaluatie van data-organisatiemethodes en data-zoekmethodes. Leuven, s.n., 1973. 3 V.
7. PENNYCUICK Roy A. (1973)
   The economics of the ecological syndrome. Leuven, Acco, 1973. XII, 177 pp.

8. KAWATA T. Bualum (1973)
   Formation du capital d'origine belge, dette publique et stratégie du développement au Zaire. Leuven, K. U. Leuven, 1973. V, 342 pp.
9. DONCKELS Rik (1974)
   Doelmatige oriëntering van de sectorale subsidiepolitiek in België: een theoretisch onderzoek met empirische toetsing. Leuven, K. U. Leuven, 1974. VII, 156 pp.
10. VERHELST Maurice (1974)
    Contribution to the analysis of organizational information systems and their financial benefits. Leuven, K. U. Leuven, 1974. 2 V.
11. CLEMEUR Hugo (1974)
    Enkele verzekeringstechnische vraagstukken in het licht van de nutstheorie. Leuven, Aurelia scientifica, 1974. 193 pp.
12. HEYVAERT Edward (1975)
    De ontwikkeling van de moderne bank- en krediettechniek tijdens de zestiende en zeventiende eeuw in Europa en te Amsterdam in het bijzonder. Leuven, K. U. Leuven, 1975. 186 pp.
13. VERTONGHEN Robert (1975)
    Investeringscriteria voor publieke investeringen: het uitwerken van een operationele theorie met een toepassing op de verkeersinfrastructuur. Leuven, Acco, 1975. 254 pp.
14. Niet toegekend.
15. VANOVERBEKE Lieven (1975)
    Microeconomisch onderzoek van de sectoriële arbeidsmobiliteit. Leuven, Acco, 1975. 205 pp.
16. DAEMS Herman (1975)
    The holding company: essays on financial intermediation, concentration and capital market imperfections in the Belgian economy. Leuven, K. U. Leuven, 1975. XII, 268 pp.
17. VAN ROMPUY Eric (1975)
    Groot-Brittannië en de Europese monetaire integratie: een onderzoek naar de gevolgen van de Britse toetreding op de geplande Europese monetaire unie. Leuven, Acco, 1975. XIII, 222 pp.
18. MOESEN Wim (1975)
    Het beheer van de staatsschuld en de termijnstructuur van de intrestvoeten met een toepassing voor België. Leuven, Vander, 1975. XVI, 250 pp.
19. LAMBRECHT Marc (1976)
    Capacity constrained multi-facility dynamic lot-size problem. Leuven, K. U. Leuven, 1976. 165 pp.

20. RAYMAECKERS Erik (1976)
De mens in de onderneming en de theorie van het producenten-gedrag: een bijdrage tot transdisciplinaire analyse. Leuven, Acco, 1976. XIII, 538 pp.

21. TEJANO Albert (1976)
Econometric and input-output models in development planning: the case of the Philippines. Leuven, K. U. Leuven, 1976. XX, 297 pp.

22. MARTENS Bernard (1977)
Prijsbeleid en inflatie met een toepassing op België. Leuven, K. U. Leuven, 1977. IV, 253 pp.

23. VERHEIRSTRAETEN Albert (1977)
Geld, krediet en intrest in de Belgische financiële sector. Leuven, Acco, 1977. XXII, 377 pp.

24. GHEYSSENS Lieven (1977)
International diversification through the government bond market: a risk-return analysis. Leuven, s.n., 1977. 188 pp.

25. LEFEBVRE Chris (1977)
Boekhoudkundige verwerking en financiële verslaggeving van huurkoop-transacties en verkopen op afbetaling bij ondernemingen die aan consumenten verkopen. Leuven, K. U. Leuven, 1977. 228 pp.

26. KESENNE Stefan (1978)
Tijdsallocatie en vrijetijdsbesteding: een econometrisch onderzoek. Leuven, s.n., 1978. 163 pp.

27. VAN HERCK Gustaaf (1978)
Aspecten van optimaal bedrijfsbeleid volgens het marktwaardecriterium: een risico-rendements-analyse. Leuven, K. U. Leuven, 1978. IV, 163 pp.

28. VAN POECK Andre (1979)
World price trends and price and wage development in Belgium: an investigation into the relevance of the Scandinavian model of inflation for Belgium. Leuven, s.n., 1979. XIV, 260 pp.

29. VOS Herman (1978)
De industriële technologieverwerving in Brazilië: een analyse. Leuven, s.n., 1978. onregelmatig gepagineerd.

30. DOMBRECHT Michel (1979)
Financial markets, employment and prices in open economies. Leuven, K. U. Leuven, 1979. 182 pp.

31. DE PRIL Nelson (1979)
Bijdrage tot de actuariële studie van het bonus-malussysteem. Brussel, OAB, 1979. 112 pp.

32. CARRIN Guy (1979)
Economic aspects of social security: a public economics approach. Leuven, K. U. Leuven, 1979. onregelmatig gepagineerd

33. REGIDOR Baldomero (1979)
An empirical investigation of the distribution of stock-market prices and weak-form efficiency of the Brussels stock exchange. Leuven, K. U. Leuven, 1979. 214 pp.

34. DE GROOT Roger (1979)
Ongelijkheden voor stop loss premies gebaseerd op E.T. systemen in het kader van de veralgemeende convexe analyse. Leuven, K. U. Leuven, 1979. 155 pp.

35. CEYSSENS Martin (1979)
On the peak load problem in the presence of rationizing by waiting. Leuven, K. U. Leuven, 1979. IX, 217 pp.

36. ABDUL RAZK Abdul (1979)
Mixed enterprise in Malaysia: the case study of joint venture between Malysian public corporations and foreign enterprises. Leuven, K. U. Leuven, 1979. 324 pp.

37. DE BRUYNE Guido (1980)
Coordination of economic policy: a game-theoretic approach. Leuven, K. U. Leuven, 1980. 106 pp.

38. KELLES Gerard (1980)
Demand, supply, price change and trading volume on financial markets of the matching-order type. Vraag, aanbod, koersontwikkeling en omzet op financiële markten van het Europese type. Leuven, K. U. Leuven, 1980. 222 pp.

39. VAN EECKHOUDT Marc (1980)
De invloed van de looptijd, de coupon en de verwachte inflatie op het opbrengstverloop van vastrentende finaciële activa. Leuven, K. U. Leuven, 1980. 294 pp.

40. SERCU Piet (1981)
Mean-variance asset pricing with deviations from purchasing power parity. Leuven, s.n., 1981. XIV, 273 pp.

41. DEQUAE Marie-Gemma (1981)
Inflatie, belastingsysteem en waarde van de onderneming. Leuven, K. U. Leuven, 1981. 436 pp.

42. BRENNAN John (1982)
An empirical investigation of Belgian price regulation by prior notification: 1975 - 1979 - 1982. Leuven, K. U. Leuven, 1982. XIII, 386 pp.

43. COLLA Annie (1982)
Een econometrische analyse van ziekenhuiszorgen. Leuven, K. U. Leuven, 1982. 319 pp.

44. Niet toegekend.

45. SCHOKKAERT Eric (1982)
Modelling consumer preference formation. Leuven, K. U. Leuven, 1982. VIII, 287 pp.

46. DEGADT Jan (1982)
Specificatie van een econometrisch model voor vervuilingsproblemen met proeven van toepassing op de waterverontreiniging in België. Leuven, s.n., 1982. 2 V.

47. LANJONG Mohammad Nasir (1983)
A study of market efficiency and risk-return relationships in the Malaysian capital market. s.l., s.n., 1983. XVI, 287 pp.

48. PROOST Stef (1983)
De allocatie van lokale publieke goederen in een economie met een centrale overheid en lokale overheden. Leuven, s.n., 1983. onregelmatig gepagineerd.

49. VAN HULLE Cynthia (1983)
Shareholders' unanimity and optimal corporate decision making in imperfect capital markets. s.l., s.n., 1983. 147 pp. + appendix.

50. VAN WOUWE Martine (2/12/83)
Ordening van risico's met toepassing op de berekening van ultieme ruïnekansen. Leuven, s.n., 1983. 109 pp.

51. D'ALCANTARA Gonzague (15/12/83)
SERENA: a macroeconomic sectoral regional and national account econometric model for the Belgian economy. Leuven, K. U. Leuven, 1983. 595 pp.

52. D'HAVE Piet (24/02/84)
De vraag naar geld in België. Leuven, K. U. Leuven, 1984. XI, 318 pp.

53. MAES Ivo (16/03/84)
The contribution of J.R. Hicks to macro-economic and monetary theory. Leuven, K. U. Leuven, 1984. V, 224 pp.

54. SUBIANTO Bambang (13/09/84)
A study of the effects of specific taxes and subsidies on a firms' R&D investment plan. s.l., s.n., 1984. V, 284 pp.

55. SLEUWAEGEN Leo (26/10/84)
Location and investment decisions by multinational enterprises in Belgium and Europe. Leuven, K. U. Leuven, 1984. XII, 247 pp.

56. GEYSKENS Erik (27/03/85)
Produktietheorie en dualiteit. Leuven, s.n., 1985. VII, 392 pp.

57. COLE Frank (26/06/85)
Some algorithms for geometric programming. Leuven, K. U. Leuven, 1985. 166 pp.

58. STANDAERT Stan (26/09/86)
A study in the economics of repressed consumption. Leuven, K. U. Leuven, 1986. X, 380 pp.

59. DELBEKE Jos (03/11/86)
Trendperioden in de geldhoeveelheid van België 1877-1983: een theoretische en empirische analyse van de "Banking school" hypothese. Leuven, K. U. Leuven, 1986. XII, 430 pp.

60. VANTHIENEN Jan (08/12/86)
Automatiseringsaspecten van de specificatie, constructie en manipulatie van beslissingstabellen. Leuven, s.n., 1986. XIV, 378 pp.

61. LUYTEN Robert (30/04/87)
A systems-based approach for multi-echelon production/inventory systems. s.l., s.n., 1987. 3V.

62. MERCKEN Roger (27/04/87)
De invloed van de data base benadering op de interne controle. Leuven, s.n., 1987. XIII, 346 pp.

63. VAN CAYSEELE Patrick (20/05/87)
Regulation and international innovative activities in the pharmaceutical industry. s.l., s.n., 1987. XI, 169 pp.

64. FRANCOIS Pierre (21/09/87)
De empirische relevantie van de independence from irrelevant alternatives. Assumptie indiscrete keuzemodellen. Leuven, s.n., 1987. IX, 379 pp.

65. DECOSTER André (23/09/88)
Family size, welfare and public policy. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1988. XIII, 444 pp.

66. HEIJNEN Bart (09/09/88)
Risicowijziging onder invloed van vrijstellingen en herverzekeringen: een theoretische analyse van optimaliteit en premiebepaling. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1988. onregelmatig gepagineerd.

67. GEEROMS Hans (14/10/88)
Belastingvermijding. Theoretische analyse van de determinanten van de belastingontduiking en de belastingontwijking met empirische verificaties. Leuven, s.n., 1988. XIII, 409, 5 pp.

68. PUT Ferdi (19/12/88)
Introducing dynamic and temporal aspects in a conceptual (database) schema. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1988. XVIII, 415 pp.

69. VAN ROMPUY Guido (13/01/89)
A supply-side approach to tax reform programs. Theory and empirical evidence for Belgium. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1989. XVI, 189, 6 pp.

70. PEETERS Ludo (19/06/89)
Een ruimtelijk evenwichtsmodel van de graanmarkten in de E.G.: empirische specificatie en beleidstoepassingen. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1989. XVI, 412 pp.

71. PACOLET Jozef (10/11/89) Marktstructuur en operationele efficiëntie in de Belgische financiële sector. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1989. XXII, 547 pp.

72. VANDEBROEK Martina (13/12/89)
Optimalisatie van verzekeringscontracten en premieberekeningsprincipes. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1989. 95 pp.

73. WILLEKENS Francois (1990)
Determinance of government growth in industrialized countries with applications to Belgium. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1990. VI, 332 pp.

74. VEUGELERS Reinhilde (02/04/90)
Scope decisions of multinational enterprises. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1990. V, 221 pp.

75. KESTELOOT Katrien (18/06/90)
Essays on performance diagnosis and tacit cooperation in international oligopolies. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1990. 227 pp.

76. WU Changqi (23/10/90)
Strategic aspects of oligopolistic vertical integration. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1990. VIII, 222 pp.

77. ZHANG Zhaoyong (08/07/91)
A disequilibrium model of China's foreign trade behaviour. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1991. XII, 256 pp.

78. DHAENE Jan (25/11/91)
Verdelingsfuncties, benaderingen en foutengrenzen van stochastische grootheden geassocieerd aan verzekeringspolissen en -portefeuilles. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1991. 146 pp.

79. BAUWELINCKX Thierry (07/01/92)
Hierarchical credibility techniques. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. 130 pp.

80. DEMEULEMEESTER Erik (23/3/92)
Optimal algorithms for various classes of multiple resource-constrained project scheduling problems. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. 180 pp.

81. STEENACKERS Anna (1/10/92)
Risk analysis with the classical actuarial risk model: theoretical extensions and applications to Reinsurance. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. 139 pp.

82. COCKX Bart (24/09/92)
The minimum income guarantee. Some views from a dynamic perspective. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. XVII, 401 pp.

83. MEYERMANS Eric (06/11/92)
Econometric allocation systems for the foreign exchange market: Specification, estimation and testing of transmission mechanisms under currency substitution. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. XVIII, 343 pp.

84. CHEN Guoqing (04/12/92)
Design of fuzzy relational databases based on fuzzy functional dependency. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. 176 pp.

85. CLAEYS Christel (18/02/93)
Vertical and horizontal category structures in consumer decision making: The nature of product hierarchies and the effect of brand typicality. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 348 pp.

86. CHEN Shaoxiang (25/03/93)
The optimal monitoring policies for some stochastic and dynamic production processes. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 170 pp.

87. OVERWEG Dirk (23/04/93)
Approximate parametric analysis and study of cost capacity management

of computer configurations. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 270 pp.

88. DEWACHTER Hans (22/06/93)
Nonlinearities in speculative prices: The existence and persistence of nonlinearity in foreign exchange rates. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 151 pp.

89. LIN Liangqi (05/07/93)
Economic determinants of voluntary accounting choices for R & D expenditures in Belgium. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 192 pp.

90. DHAENE Geert (09/07/93)
Encompassing: formulation, properties and testing. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 117 pp.

91. LAGAE Wim (20/09/93)
Marktconforme verlichting van soevereine buitenlandse schuld door private crediteuren: een neo-institutionele analyse. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 241 pp.

92. VAN DE GAER Dirk (27/09/93)
Equality of opportunity and investment in human capital. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 172 pp.

93. SCHROYEN Alfred (28/02/94)
Essays on redistributive taxation when monitoring is costly. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 203 pp. + V.

94. STEURS Geert (15/07/94)
Spillovers and cooperation in research and development. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 266 pp.

95. BARAS Johan (15/09/94)
The small sample distribution of the Wald, Lagrange multiplier and likelihood ratio tests for homogeneity and symmetry in demand analysis: a Monte Carlo study. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 169 pp.

96. GAEREMYNCK Ann (08/09/94)
The use of depreciation in accounting as a signalling device. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 232 pp.

97. BETTENDORF Leon (22/09/94)
A dynamic applied general equilibrium model for a small open economy. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 149 pp.

98. TEUNEN Marleen (10/11/94)
Evaluation of interest randomness in actuarial quantities. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 214 pp.

99. VAN OOTEGEM Luc (17/01/95)
An economic theory of private donations. Leuven. K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 236 pp.

100. DE SCHEPPER Ann (20/03/95)
Stochastic interest rates and the probabilistic behaviour of actuarial functions. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 211 pp.

101. LAUWERS Luc (13/06/95)
Social choice with infinite populations. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 79 pp.

102. WU Guang (27/06/95)
A systematic approach to object-oriented business modeling. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 248 pp.

103. WU Xueping (21/08/95)
Term structures in the Belgian market: model estimation and pricing error analysis. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 133 pp.

104. PEPERMANS Guido (30/08/95)
Four essays on retirement from the labor force. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 128 pp.

105. ALGOED Koen (11/09/95)
Essays on insurance: a view from a dynamic perspective. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 136 pp.

106. DEGRYSE Hans (10/10/95)
Essays on financial intermediation, product differentiation, and market structure. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 218 pp.

107. MEIR Jos (05/12/95)
Het strategisch groepsconcept toegepast op de Belgische financiële sector.

Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 257 pp.

108. WIJAYA Miryam Lilian (08/01/96)
Voluntary reciprocity as an informal social insurance mechanism: a game theoretic approach. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 124 pp.

109. VANDAELE Nico (12/02/96)
The impact of lot sizing on queueing delays: multi product, multi machine models. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 243 pp.

110. GIELENS Geert (27/02/96)
Some essays on discrete time target zones and their tails. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 131 pp.

111. GUILLAUME Dominique (20/03/96)
Chaos, randomness and order in the foreign exchange markets. Essays on the modelling of the markets. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 171 pp.

112. DEWIT Gerda (03/06/96)
Essays on export insurance subsidization. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 186 pp.

113. VAN DEN ACKER Carine (08/07/96)
Belief-function theory and its application to the modeling of uncertainty in financial statement auditing. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 147 pp.

114. IMAM Mahmood Osman (31/07/96)
Choice of IPO Flotation Methods in Belgium in an Asymmetric Information Framework and Pricing of IPO's in the Long-Run. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 221 pp.

115. NICAISE Ides (06/09/96)
Poverty and Human Capital. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 209 pp.

116. EYCKMANS Johan (18/09/97)
On the Incentives of Nations to Join International Environmental Agreements. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1997. XV + 348 pp.

117. CRISOLOGO-MENDOZA Lorelei (16/10/97)
Essays on Decision Making in Rural Households: a study of three villages in the Cordillera Region of the Philippines. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1997. 256 pp.

118. DE REYCK Bert (26/01/98)
Scheduling Projects with Generalized Precedence Relations: Exact and Heuristic Procedures. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. XXIV + 337 pp.
119. VANDEMAELE Sigrid (30/04/98)
Determinants of Issue Procedure Choice within the Context of the French IPO Market: Analysis within an Asymmetric Information Framework. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. 241 pp.
120. VERGAUWEN Filip (30/04/98)
Firm Efficiency and Compensation Schemes for the Management of Innovative Activities and Knowledge Transfers. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. VIII + 175 pp.
121. LEEMANS Herlinde (29/05/98)
The Two-Class Two-Server Queueing Model with Nonpreemptive Heterogeneous Priority Structures. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. 211 pp.
122. GEYSKENS Inge (4/09/98)
Trust, Satisfaction, and Equity in Marketing Channel Relationships. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. 202 pp.
123. SWEENEY John (19/10/98)
Why Hold a Job ? The Labour Market Choice of the Low-Skilled. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. 278 pp.
124. GOEDHUYS Micheline (17/03/99)
Industrial Organisation in Developing Countries, Evidence from Côte d'Ivoire. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 251 pp.
125. POELS Geert (16/04/99)
On the Formal Aspects of the Measurement of Object-Oriented Software Specifications. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 507 pp.
126. MAYERES Inge (25/05/99)
The Control of Transport Externalities: A General Equilibrium Analysis. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. XIV + 294 pp.
127. LEMAHIEU Wilfried (5/07/99)
Improved Navigation and Maintenance through an Object-Oriented Approach to Hypermedia Modelling. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 284 pp.

128. VAN PUYENBROECK Tom (8/07/99)
Informational Aspects of Fiscal Federalism. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 192 pp.

129. VAN DEN POEL Dirk (5/08/99)
Response Modeling for Database Marketing Using Binary Classification. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 342 pp.

130. GIELENS Katrijn (27/08/99)
International Entry Decisions in the Retailing Industry: Antecedents and Performance Consequences. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 336 pp.

131. PEETERS Anneleen (16/12/99)
Labour Turnover Costs, Employment and Temporary Work. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 207 pp.

132. VANHOENACKER Jurgen (17/12/99)
Formalizing a Knowledge Management Architecture Meta-Model for Integrated Business Process Management. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 252 pp.

133. NUNES Paulo (20/03/2000)
Contingent Valuation of the Benefits of Natural Areas and its Warmglow Component. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. XXI + 282 pp.

134. VAN DEN CRUYCE Bart (7/04/2000)
Statistische discriminatie van allochtonen op jobmarkten met rigide lonen. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. XXIII + 441 pp.

135. REPKINE Alexandre (15/03/2000)
Industrial restructuring in countries of Central and Eastern Europe: Combining branch-, firm- and product-level data for a better understanding of Enterprises' behaviour during transition towards market economy. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. VI + 147 pp.

136. AKSOY, Yunus (21/06/2000)
Essays on international price rigidities and exchange rates. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. IX + 236 pp.

137. RIYANTO, Yohanes Eko (22/06/2000)
Essays on the internal and external delegation of authority in firms. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. VIII + 280 pp.

138. HUYGHEBAERT, Nancy (20/12/2000)
The Capital Structure of Business Start-ups. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. VIII + 332 pp.

139. FRANCKX Laurent (22/01/2001)
Ambient Inspections and Commitment in Environmental Enforcement. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. VIII + 286 pp.

140. VANDILLE Guy (16/02/2001)
Essays on the Impact of Income Redistribution on Trade. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. VIII + 176 pp.

141. MARQUERING Wessel (27/04/2001)
Modeling and Forecasting Stock Market Returns and Volatility. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. V + 267 pp.

142. FAGGIO Giulia (07/05/2001)
Labor Market Adjustment and Enterprise Behavior in Transition. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. 150 pp.

143. GOOS Peter (30/05/2001)
The Optimal Design of Blocked and Split-plot experiments. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. X + 224 pp.

144. LABRO Eva (01/06/2001)
Total Cost of Ownership Supplier Selection based on Activity Based Costing and Mathematical Programming. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. 217 pp.

145. VANHOUCKE Mario (07/06/2001)
Exact Algorithms for various Types of Project Scheduling Problems. Nonregular Objectives and time/cost Trade-offs. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. 316 pp.

146. BILSEN Valentijn (28/08/2001)
Entrepreneurship and Private Sector Development in Central European Transition Countries. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. XVI + 188 pp.

147. NIJS Vincent (10/08/2001)
Essays on the dynamic Category-level Impact of Price promotions. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001.

148. CHERCHYE Laurens (24/09/2001)
Topics in Non-parametric Production and Efficiency Analysis. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. VII + 169 pp.

149. VAN DENDER Kurt (15/10/2001)
Aspects of Congestion Pricing for Urban Transport. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. VII + 203 pp.

150. CAPEAU Bart (26/10/2001)
In defence of the excess demand approach to poor peasants' economic behaviour. Theory and Empirics of non-recursive agricultural household modelling. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. XIII + 286 pp.

151. CALTHROP Edward (09/11/2001)
Essays in urban transport economics. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001.

152. VANDER BAUWHEDE Heidi (03/12/2001)
Earnings management in an Non-Anglo-Saxon environment. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. 408 pp.

153. DE BACKER Koenraad (22/01/2002)
Multinational firms and industry dynamics in host countries: the case of Belgium. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. VII + 165 pp.

154. BOUWEN Jan (08/02/2002)
Transactive memory in operational workgroups. Concept elaboration and case study. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 319 pp. + appendix 102 pp.

155. VAN DEN BRANDE Inge (13/03/2002)
The psychological contract between employer and employee: a survey among Flemish employees. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. VIII + 470 pp.

156. VEESTRAETEN Dirk (19/04/2002)
Asset Price Dynamics under Announced Policy Switching. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 176 pp.

157. PEETERS Marc (16/05/2002)
One Dimensional Cutting and Packing: New Problems and Algorithms. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. IX + 247 pp.

158. SKUDELNY Frauke (21/05/2002)
Essays on The Economic Consequences of the European Monetary Union. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002.

159. DE WEERDT Joachim (07/06/2002)
Social Networks, Transfers and Insurance in Developing countries. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. VI + 129 pp.

160. TACK Lieven (25/06/2002)
Optimal Run Orders in Design of Experiments. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. XXXI + 344 pp.

161. POELMANS Stephan (10/07/2002)
Making Workflow Systems work. An investigation into the Importance of Task-appropriation fit, End-user Support and other Technological Characteristics. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 237 pp.

162. JANS Raf (26/09/2002)
Capacitated Lot Sizing Problems : New Applications, Formulations and Algorithms. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002.

163. VIAENE Stijn (25/10/2002)
Learning to Detect Fraud from enriched Insurance Claims Data (Context, Theory and Applications). Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 315 pp.

164. AYALEW Tekabe (08/11/2002)
Inequality and Capital Investment in a Subsistence Economy. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. V + 148 pp.

165. MUES Christophe (12/11/2002)
On the Use of Decision Tables and Diagrams in Knowledge Modeling and Verification. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 222 pp.

166. BROCK Ellen (13/03/2003)
The Impact of International Trade on European Labour Markets. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002.

167. VERMEULEN Frederic (29/11/2002)
Essays on the collective Approach to Household Labour Supply. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. XIV + 203 pp.

168. CLUDTS Stephan (11/12/2002)
Combining participation in decision-making with financial participation: theoretical and empirical perspectives. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. XIV + 247 pp.

169. WARZYNSKI Frederic (09/01/2003)
The dynamic effect of competition on price cost margins and innovation. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

170. VERWIMP Philip (14/01/2003)
Development and genocide in Rwanda ; a political economy analysis of peasants and power under the Habyarimana regime. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

171. BIGANO Andrea (25/02/2003)
Environmental regulation of the electricity sector in a European Market Framework. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. XX + 310 pp.

172. MAES Konstantijn (24/03/2003)
Modeling the Term Structure of Interest Rates Across Countries. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. V+246 pp.

173. VINAIMONT Tom (26/02/2003)
The performance of One- versus Two-Factor Models of the Term Structure of Interest Rates. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen. 2003.

174. OOGHE Erwin (15/04/2003)
Essays in multi-dimensional social choice. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. VIII+108 pp.

175. FORRIER Anneleen (25/04/2003)
Temporary employment, employability and training. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

176. CARDINAELS Eddy (28/04/2003)
The role of cost system accuracy in managerial decision making. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. 144 pp.

177. DE GOEIJ Peter (02/07/2003)
Modeling Time-Varying Volatility and Interest Rates. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. VII+225 pp.

178. LEUS Roel (19/09/2003)
The generation of stable project plans. Complexity and exact algorithms. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

179. MARINHEIRO Carlos (23/09/2003)
EMU and fiscal stabilisation policy: the case of small countries. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

180. BAESSENS Bart (24/09/2003)
Developing intelligent systems for credit scoring using machine learning techniques. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

181. KOCZY Laszlo (18/09/2003)
Solution concepts and outsider behaviour in coalition formation games. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

182. ALTOMONTE Carlo (25/09/2003)
Essays on Foreign Direct Investment in transition countries: learning from the evidence. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

183. DRIES Liesbeth (10/11/2003)
Transition, Globalisation and Sectoral Restructuring: Theory and Evidence from the Polish Agri-Food Sector. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

184. DEVOOGHT Kurt (18/11/2003)
Essays On Responsibility-Sensitive Egalitarianism and the Measurement of Income Inequality. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

185. DELEERSNYDER Barbara (28/11/2003)
Marketing in Turbulent Times. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

186. ALI Daniel (19/12/2003)
Essays on Household Consumption and Production Decisions under Uncertainty in Rural Ethiopia. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

187. WILLEMS Bert (14/01/2004)
Electricity networks and generation market power. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

188. JANSSENS Gust (30/01/2004)
Advanced Modelling of Conditional Volatility and Correlation in Financial Markets. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

189. THOEN Vincent (19/01/2004)
On the valuation and disclosure practices implemented by venture capital providers. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

190. MARTENS Jurgen (16/02/2004)
A fuzzy set and stochastic system theoretic technique to validate simulation models. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

191. ALTAVILLA Carlo (21/05/2004)
Monetary policy implementation and transmission mechanisms in the Euro area. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

192. DE BRUYNE Karolien (07/06/2004)
Essays in the location of economic activity. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

193. ADEM Jan (25/06/2004)
Mathematical programming approaches for the supervised classification problem. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

194. LEROUGE Davy (08/07/2004)
Predicting Product Preferences: the effect of internal and external cues. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

195. VANDENBROECK Katleen (16/07/2004)
Essays on output growth, social learning and land allocation in agriculture : micro-evidence from Ethiopia and Tanzania. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

196. GRIMALDI Maria (03/09/2004)
The exchange rate, heterogeneity of agents and bounded rationality. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

197. SMEDTS Kristien (26/10/2004)
Financial integration in EMU in the framework of the no-arbitrage theory. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

198. KOEVOETS Wim (12/11/2004)
Essays on Unions, Wages and Employment. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

199. CALLENS Marc (22/11/2004)
Essays on multilevel logistic Regression. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

200. RUGGOO Arvind (13/12/2004)
Two stage designs robust to model uncertainty. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

201. HOORELBEKE Dirk (28/01/2005)
Bootstrap and Pivoting Techniques for Testing Multiple Hypotheses. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

202. ROUSSEAU Sandra (17/02/2005)
Selecting Environmental Policy Instruments in the Presence of Incomplete Compiance. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

203. VAN DER MEULEN Sofie (17/02/2005)
Quality of Financial Statements: Impact of the external auditor and applied accounting standards. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

204. DIMOVA Ralitza (21/02/2005)
Winners and Losers during Structural Reform and Crisis: the Bulgarian Labour Market Perspective. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

205. DARKIEWICZ Grzegorz (28/02/2005)
Value-at-risk in Insurance and Finance: the Comonotonicity Approach. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

206. DE MOOR Lieven (20/05/2005)
The Structure of International Stock Returns: Size, Country and Sector Effects in Capital Asset Pricing. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

207. EVERAERT Greetje (27/06/2005)
Soft Budget Constraints and Trade Policies: The Role of Institutional and External Constraints. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

208. SIMON Steven (06/07/2005)
The Modeling and Valuation of complex Derivatives: The Impact of the Choice of the term structure model. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

209. MOONEN Linda (23/09/2005)
Algorithms for some Graph-Theoretical Optimization Problems. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

210. COUCKE Kristien (21/09/2005)
Firm and industry adjustment under de-industrialisation and globalization of the Belgian economy. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

211. DECAMPS Marc (21/10/2005)
Some actuarial and financial applications of generalized diffusion processes. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

212. KIM Helena (29/11/2005)
Escalation games: an instrument to analyze conflicts. The strategic approach to the bargaining problem. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

213. GERMENJI Etleva (06/01/2006)
Essays on the Economics of Emigration from Albania. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

214. BELIEN Jeroen (18/01/2006)
Exact and Heuristic Methodologies for Scheduling in Hospitals: Problems, Formulations and Algorithms. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

215. JOOSSENS Kristel (10/02/2006)
Robust discriminant analysis. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

216. VRANKEN LIESBET (13/02/2006)
Land markets and production efficiency in transition economies. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

217. VANSTEENKISTE ISABEL (22/02/2006)
Essays on non-linear modelling in international macroeconomics. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

218. WUYTS Gunther (31/03/2006)
Essays on the liquidity of financial markets. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

219. DE BLANDER Rembert (28/04/2006)
Essays on endogeneity and parameter heterogeneity in cross-section and

panel data. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

220. DE LOECKER Jan (12/05/2006)
Industry dynamics and productivity. Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

221. LEMMENS Aurélie (12/05/2006)
Advanced classification and time-series methods in marketing, Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.