Multivariate Extensions to Discrete Choice Modeling

Multivariate Extensions to Discrete Choice Modeling

Multivariate uitbreidingen van discrete keuzemodellen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on

Thursday, November 12, 2015 at 09:30 hours

by

KOEN BEL

born in Voorhout, The Netherlands

Erasmus University Rotterdam

Doctorate Committee

Promotor: Prof.dr. R. Paap

Other members: Prof.dr. D. Fok

Prof.dr. Ph.H.B.F Franses

Prof.dr. C. Croux

ISBN: 978 90 5170 653 6

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 629 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Acknowledgments

Following the initial title of my PhD proposal, *Forecasting in a Changing World*, I could not have forecasted the thesis to look like it does right now. The world of a PhD candidate highly changes in four years. Colleagues come and go, skills improve and interests change. The topic of the thesis has changed from macro-economic forecasts to micro-economic estimation methods. Apparently, the world of a PhD is too volatile to accurately forecast its future.

The four years as a PhD candidate brought me a lot. Visiting conferences in London, Toulouse, Pisa and Montréal improved my presentation skills. Creating papers improved my writing skills and my command of English. Technical discussions with my supervisor and other colleagues improved my knowledge of the field of econometrics. That is, the changing world of the PhD candidate clearly has an upward trend.

I could not have reached all of this without the help of many. Richard has been of great support, not only as my supervisor but also as sounding board. His enthusiasm on econometric research inspired me. Richard has done much more than he ought to as PhD supervisor, of which I am grateful. I further thank Dennis and Pieter for their suggestions and work as co-authors of two of the chapters in this thesis. Furhtermore, the committee consisting of Dennis Fok, Philip Hans Franses and Christophe Croux gave useful comments which helped to improve the thesis.

I further thank the Econometric Institute and Tinbergen Institute to give me the opportunity to work for four years with this booklet as a result. Both secretaries made the four years much easier. Special thanks to Anneke Kop, who answered my recurrent questions with endurance.

The fruitful thoughts during lunch with colleagues helped a lot, no matter whether it was on econometrics, politics, ideology or whatever came to mind. Bart, Bert, Bruno, Dennis, Didier, Francine, Myrthe, Tom and Victor helped to offset against the daily practice of a PhD candidate. Having colleagues who are as well friends results in a very pleasant working environment. I further thank Gertjan, Pieter and Ronald, Bart, Francine and Tom and Didier for their company during conference visits. Finally and most importantly, I thank my roommates Bert and Mathijn for their patience with me. Bert, your listening ear and willingness to think along with me has been of great importance to me. The time you have devoted to me is worth more than a PhD-salary.

Following Monty Phytons *"And now for something completely different"*, the offset against work has been of great importance to me. Elwin, Guus, Iris, Jan, Jordy, Marloes, Roy, Sjoerd and Stefan: the Friday nights and other activities are great and I hope there are still many to come. My roomies made my stay in Rotterdam very pleasant. Finally, I would like to thank my friends from studies for our occasional meetings.

I thank my paranymphs Anneke and Didier to be on my side for one hour. Your support in the organisation of the defense has been of great help. You are allowed to sit.

In the end, it all comes back to family. Mom and dad, you supported me all the way through although you have no clue what I am doing. Your tips and comments on life have been of great help. Finally, Ilse: you gave my life an extra dimension. Your presence gives me rest and pleasure.

Contents

1 Introduction **1**

1.1 Discrete Choice Modeling 1

1.2 Modeling the Impact of Forecasts 4

2 Parameter Estimation in Multivariate Logit Models **7**

2.1 Introduction 7

2.2 Model Specification 9

2.3 Parameter Inference 11

2.4 Monte Carlo Study 16

2.5 Application 23

2.6 Conclusion 27

3 A Multivariate Model for Multinomial Choices **29**

3.1 Introduction 29

3.2 Model Specification 31

3.2.1 A Bivariate Trinomial Logit Model 34

3.2.2 Parameter Inference 35

3.2.3 Monte Carlo Study 37

3.3 A Panel Specification 41

3.3.1 Parameter Inference 42

3.4 Application 43

3.4.1 Survey Data on Satisfaction 44

3.4.2 Household Panel Scanner Data 47

3.5 Conclusion 50

3.A Derivations 51

3.A.1 Joint probabilities in MV-MNL 51
3.A.2 Choice probability conditional on sufficient statistic 52
3.A.3 Composite Conditional Likelihood in panel data setting 53

4 The Multivariate Dale Model 55
4.1 Introduction 55
4.2 The Bivariate Dale Model 56
4.3 The Multivariate Dale Model and Composite Likelihood Estimation 58
4.3.1 Parameter Inference 60
4.4 Simulation Study 62
4.5 Application 65
4.6 Conclusion 69
4.A Plackett Distribution 70
4.B Derivatives 71
4.C Tables 72

5 Covariates in Two-Stage Choice Models 75
5.1 Introduction 75
5.2 Model Specification 76
5.3 Simulation Study 78
5.3.1 Simulation Results 80
5.4 Conclusion 87

6 Modeling the Impact of Forecasts 89
6.1 Introduction 89
6.2 Model Specification 90
6.2.1 Specification of the influential forecast 91
6.2.2 Specification of the time-varying threshold 92
6.3 Statistical Inference 93
6.3.1 Estimation procedure 93
6.3.2 Nonlinearity test 94
6.4 Application 98
6.4.1 Model specification 98

6.4.2 Model selection 99
6.4.3 Parameter interpretation 101
6.5 Conclusion 109
6.A Tables 111

Nederlandse samenvatting (Summary in Dutch) 113

Bibliography 117

Chapter 1

Introduction

This thesis on econometric modeling consists of two parts. First, I consider multivariate extensions to univariate discrete choice models. The focus of the chapters is on model representation and parameter estimation methods. Univariate choice models are extended to allow for associations between different univariate choices. When the number of choices involved is large, parameter estimation in these multivariate extensions leads to numerical problems. Therefore, new feasible parameter estimation methods are introduced. In the second part, I discuss the effect of forecasts on future values of macro-economic variables. Since economic agents react to forecasts, these might affect the course of the economy. Below, I provide a short introduction to both parts. Although the chapters in this thesis are related, they can be read independently.

1.1 Discrete Choice Modeling

Each individual makes numerous discrete choices a day. For example, traveling to work can be done by bike, by car or by public transport. In a supermarket, individuals choose between several brands. The choices people make are of interest in many field of research, such as marketing, micro-economics and transportation.

Choice models aim to investigate how these discrete decisions are made and what aspects influence the choices. For example, prices, personal income and advertisements may influence brand choice. Information from choice models provides insights in choice processes and helps, for example, marketeers to decide upon future pricing and promotion.

Econometric choice modeling originated in 1860 and evolved further with pioneering work of Nobel Prize winner Daniel McFadden in the 1970s. Since then, a vast literature on the decision process has emerged. Ben-Akiva and Lerman (1985), Train (2003) and Agresti (2007) provide good overviews of the practice of choice modeling. The four chapters in the first part of this thesis extend the choice literature in several ways. The first three chapters discuss multivariate discrete choices (stemming from Cox, 1972). Multivariate extensions are rare since univariate discrete choice models are not easily scalable to higher dimensions. A large part of the discussion in these chapters is devoted to deriving feasible parameter estimation methods which can easily be scaled to high choice dimensions. The fourth chapter deals with a two-stage decision process (originating with Swait, 1984) and discusses the division of covariates over these stages. Below I provide a summary of the contributions of these chapters.

Chapter 2 is based on Bel et al. (2014) and involves multivariate yes/no-decisions. Discrete yes/no-decisions may give information about other choices that have been made. For example, if I decide to visit an outlet shop (*yes*) I will not likely visit a shop with expensive brands (*no*) during the same trip. That is, discrete choices are interrelated and Chapter 2 investigates these associations using multivariate binary choice models.

Cox (1972) describes these multivariate choices. The paper gives a wide range of model specifications, from a restrictive independence model to a full multinomial representation over the complete set of multivariate binary choices. Since both do not incorporate the association structure between decisions, Cox (1972) introduces a logistic model specification. This specification is the basis of Chapters 2 to 4 in this thesis. Russell and Petersen (2000) show that combining conditional choice models using Besag (1974) leads to the multivariate logistic specification of Cox (1972).

The focus of this chapter is parameter estimation when the number of correlated binary choices is large. Cox (1972) states that difficulties arise when the number of choices increase. Russell and Petersen (2000) state that "as the number of categories becomes large, the approach taken [. . .] will clearly become infeasible". Chapter 2 shows that these problems are circumvented by by introducing three scalable estimation methods, which result in accurate parameter estimates of the multivariate choice model even when many correlated choices are involved. The three methods are: stratified importance sampling, a generalized method of moments and composite conditional likelihood (Lindsay, 1988). The performance of these methods is analyzed using several simulation studies. In terms of accuracy, efficiency and speed, composite likelihood turns out to be the best alternative for full likelihood estimation. Although the composite likelihood estimator is inefficient compared to a regular maximum likelihood approach, the increase in estimation speed certainly outweighs the loss in efficiency for the problem at hand.

Because of its accuracy and efficiency, the composite likelihood method is also applied in Chapters 3 and 4. These chapters extend on the ideas of Chapter 2. Chapter 3 is based on Bel and Paap (2014) and extends the multivariate binary logit of Cox (1972) to a multivariate multinomial logit model for estimating associations between multinomial choices. For example, if I choose the cheap alternative for product A, I am probably also likely to choose the cheap alternative for product B. The chapter contributes to the literature by introducing a new and easily scalable model specification for multivariate multinomial choices. Where existing approaches cannot handle large dimensions, I show that the combination of the model specification and the use of the composite likelihood parameter estimation method is feasible in large multivariate problems.

Chapter 4 is based on Bel and Schoonees (2015) and involves multivariate ordered choices. Choice options are specifically ordered as in Likert scale survey questions. I advocate a multivariate extension of the existing bivariate ordered choice Dale (1986) model. The Dale (1986) model is a global odds ratio model where the correlations between two univariate choices are described by so-called global odds ratios. This model specification results in a Plackett (1965) distribution function. Unfortunately, this distribution function is not easily scalable to higher dimensions where more than two ordered choices are made. That is, Molenberghs and Lesaffre (1994) show that the Plackett distribution does not have an analytical expression if the dimension is larger than three. This highly complicates parameter estimation. Chapter 4 extends the Dale model to a multivariate setting by again using a composite likelihood method for parameter estimation. Since an expression for the bivariate distribution function is available from Dale (1986), I use composite pairwise likelihood. Simulation studies show that the approach leads to accurate parameter estimates and that efficiency losses are rather small.

Chapter 5 is the final chapter on discrete choice modeling and is based on Bel and Paap

(2015). This chapter focuses on modeling the situation where individual do not consider all available choice options when making their final choice. That is, individuals take decisions in two stages: first, an (unobserved) consideration set is formed and second, the final choice is made from this subset of choice items.

The idea of two stages in the choice process originates with Howard and Sheth (1969) and economic theory is provided by Wright and Barbour (1977). An important contribution to this literature is Swait (1984). Chapter 5 adds a critical note. Since the choice model consists of two stages and the first stage is unobservable for the econometrician, it is unclear which covariates should enter which stage of the choice process. For example, does pricing affect the consideration stage or does it directly affect choice? Chapter 5 investigates whether the role of the covariates can be determined by purely statistical measures. Simulation studies show that inference and interpretation of the unobserved consideration stage is only correct if the division of covariates over the two stages is in accordance with the actual data generating process. Statistical tests are not helpful in deciding the role of the covariates.

In sum, the novel contributions of these chapters to the discrete choice literature are (i) extensions of univariate discrete choice models to multivariate choices; (ii) computational feasible parameter estimation methods for multivariate choice models and; (iii) a critical note on the role of covariates in two-stage choice modeling.

1.2 Modeling the Impact of Forecasts

The second part of this thesis consists of one chapter based on Bel and Paap (2013) and handles the effect of forecasts on macro-economic variables. Economic agents are interested in key economic variables such as growth, inflation, interest rates and exchange rates. Since these are important features for decisions on, for example, savings and expenditures and buying or selling stocks, economic agents are highly interested in forecasts of these key variables.

I advocate that forecasts themselves have impact on the economy, since economic agents base decisions on these forecasts. For example, economic agents react if the Dutch Bureau for Economic Policy Analysis posts an extreme forecast. For time series modeling, the change in behavior of agents due to forecasts may lead to structural changes in the model parameters.

The contribution of Chapter 6 is a time series model specification which incorporates the reactions to forecasts. I use smooth transition autoregressive models as introduced by Teräsvirta and Anderson (1992). This nonlinear model handles discontinuities in the time series of interest. The smooth transition autoregressive model allows for gradual changes from regime to regime, see van Dijk (1999) for a clear overview of all aspects concerned with this type of model specifications.

Since I expect that forecasts affect the regime switching process, I use the forecasts as transition variable. This forecast can either be an exogenous expert opinion or an endogenous forecast generated by the model. For the latter, I use Dueker et al. (2007). They propose a contemporaneous transition model where the regimes are not predetermined. This thesis chapter provides a justification and interpretation of this contemporaneous model by relating it to the forecast of the time series of interest.

Results of an application to US inflation show indeed that forecasts have an impact on the level of inflation and reactions to forecasts indicate that economic agents have mean reverting behavior.

Chapter 2

Parameter Estimation in Multivariate Logit Models with Many Binary Choices

2.1 Introduction

Multivariate choice models are widely used to describe correlated binary decision data in different fields of applied research. For example, grocery product choices by consumers are likely to be correlated across different brands or product categories (Chib et al., 2002). Choices for different types of insurances are correlated (Donkers et al., 2007), and effects of a medicine treatment on two or more physiological systems are also related (Ashford and Sowden, 1970). As a final example, Feddag (2013) investigates several 'health-related quality of life'-questions in a survey among cancer patients and the answers to these questions are likely to be correlated. Hence, simultaneous discrete decisions occur in many different fields of research.

The number of choices to be made in multivariate decision problems can be rather large. The number of products in a supermarket is large; individuals have to decide upon life, car, house insurances, and so forth; and the number of questions in a survey might also be large. There is therefore a need for a model that is applicable in these settings. In principle such models are available. However, current econometric estimation methods for multivariate choice models suffer from a computational burden if the number of choices grows large.

The standard econometric model to describe correlated multivariate binary choices is the multivariate probit model (Ashford and Sowden, 1970; Edwards and Allenby, 2003).

The main disadvantage of this model is that the computation of the choice probabilities involves high-dimensional integrals which cannot be solved analytically. Numerical integration methods are not very accurate and slow in high dimensions and simulation-based estimation methods are often used instead (Cappellari and Jenkins, 2006). However, the computational efforts to perform simulation-based estimation become excessive when a large number of correlated choices is considered. To avoid the evaluation of integrals one may opt for multivariate binary decision models based on correlated logistic regressions. These models are nonetheless difficult to generalize to higher dimensions (Carey et al., 1993; Glonek and McCullagh, 1995).

To avoid these difficulties we opt for the multivariate logit [MVL] model (Cox, 1972). Russell and Petersen (2000) show that this model can be written as a restricted multinomial logit [MNL] specification over all possible outcomes of the multivariate binary choices. The multivariate choice problem over K choices is reformulated as a *multinomial* choice model over 2^K alternatives.

The problem of this MVL specification is that the outcome space of the multivariate binary random variable, and thereby the computation time, increases exponentially with the number of choices. From a practical point of view, standard maximum likelihood [ML] parameter estimation becomes computationally infeasible even for a moderate number of choices. Further, numerical problems can occur as probabilities get too small for practical use. Russell and Petersen (2000) apply the model to four binary choices only and state that "as the number of categories becomes large, the approach taken in our research will clearly become infeasible". Guimares et al. (2003) propose to use a more feasible approach based on Poisson regression. Unfortunately, this method only holds for the conditional logit specification where explanatory variables differ across choices. It therefore does not solve the infeasibility for all multivariate logit specifications.

In this chapter, we propose three novel estimation methods for the MVL model which provide parameter estimates in an acceptable amount of time even if the number of binary choices is large. In the first proposed method, we use a sampling method to reduce the number of alternatives in the estimation routine. Using the method proposed by Ben-Akiva and Lerman (1985), we can still obtain consistent estimators for the model parameters. In the second method we take advantage of the fact that the MVL model has simple conditional probabilities. We use these conditional probabilities in a composite conditional likelihood

[CCL] approach (Lindsay, 1988). In case of K binary choices, only K conditional probabilities have to be evaluated instead of 2^K joint probabilities, which reduces computing time. Furthermore, this method solves the problem of very small joint probabilities as these probabilities are not used within the estimation routine. Finally, we consider a generalized method of moments [GMM] estimator based on the conditional probabilities and hence this approach has the same advantages as the CCL approach. Monte Carlo results show that the three novel estimation methods are much faster, have similar small-sample biases as the standard ML approach of Russell and Petersen (2000), and that the loss in efficiency is very limited.

The remainder of this chapter is organized as follows. In Section 2.2 we describe the multivariate logit model as discussed by Russell and Petersen (2000). Parameter inference is considered in Section 2.3. We first present standard ML parameter estimation followed by our three alternative methods. Section 2.4 describes the results of the Monte Carlo study which compares the estimation methods with respect to computation time, small-sample bias, and efficiency. Section 2.5 gives an illustration of the MVL model for a case with 10 binary choices for store visits of households in a shopping mall. Finally, Section 2.6 concludes.

2.2 Model Specification

In this section we discuss the model specification for the multivariate logit model. We use the specification as introduced by Cox (1972) and further implemented by Russell and Petersen (2000).

Following Russell and Petersen (2000), we let Y_i denote the K-dimensional random variable describing the joint set of choices for individual $i = 1, \ldots, N$, defined as $Y_i = \{Y_{i1}, \ldots, Y_{iK}\}$, where Y_{ik} denotes the k-th binary choice for individual i, for $k = 1, \ldots, K$. The set of possible realizations of Y_i is called S which contains 2^K elements. It can immediately be seen that the number of possible realizations grows exponentially with the number of binary choices K.

The choices in Y_i may be correlated. To describe these dependencies Russell and Petersen (2000) specify the conditional probabilities of the k-th random variable Y_{ik} given all other choices, that is, y_{il} for $l \neq k$. These conditional probabilities are a logit function of individual characteristics X_i, model parameters α, β and ψ, and y_{il}, that is

$$\Pr[Y_{ik} = 1 | y_{i1}, \ldots, y_{ik-1}, y_{ik+1}, \ldots, y_{iK}, X_i] = \frac{\exp(Z_{ik})}{1 + \exp(Z_{ik})} \tag{2.1}$$

with

$$Z_{ik} = \alpha_k + X_i \beta_k + \sum_{l \neq k} y_{il} \psi_{kl}, \tag{2.2}$$

where y_{il} is the realization of Y_{il}, α_k are alternative-specific intercepts, X_i is a $(1 \times p)$-vector of explanatory variables with corresponding parameter vector β_k, and where ψ_{kl} are association parameters. The association parameters capture the correlation between Y_{ik} and Y_{il} for $l \neq k$. Positive association implies that options k and l tend to have similar values and negative association implies that they tend to be different. Conditional independence between Y_{ik} and Y_{il} occurs when $\psi_{kl} = 0$. As we can only consider correlations and no causal impacts, we have to impose $\psi_{kl} = \psi_{lk}$ for symmetry, see also Russell and Petersen (2000). The model can be extended by including explanatory variables that differ across individuals and the different binary choices. Such an extension is straightforward, but to simplify notation we do not include such variables here.

Using the results in Besag (1974), the joint distribution of Y_i follows directly from the full set of conditional distributions. Russell and Petersen (2000) show that the conditional distributions in (2.1) imply an MNL specification for the joint distribution of Y_i, that is

$$\Pr[Y_i = y_i | X_i] = \frac{\exp(\mu_{y_i})}{\sum_{s_i \in S} \exp(\mu_{s_i})}, \tag{2.3}$$

where y_i is a possible realization from the outcome space S and where μ_{y_i} is defined as

$$\mu_{y_i} = \sum_{k=1}^{K} \left(y_{ik}(\alpha_k + X_i \beta_k) + \sum_{l > k} y_{ik} y_{il} \psi_{kl} \right). \tag{2.4}$$

Hence, the parameters α_k and β_k only occur in the numerator of the probability function for $y_{ik} = 1$. Further, the association parameter ψ_{kl} only occurs in the numerator when both $y_{ik} = 1$ and $y_{il} = 1$. Note that this implies that all pairs should occur in the available data to be able to estimate these association parameters.

The interpretation of the impact of the intercept parameters and X_i follows from the log odds ratio

$$\log\left(\frac{\Pr[Y_i = y_i|X_i]}{\Pr[Y_i = (0,\ldots,0)|X_i]}\right) = \sum_{k=1}^{K} y_{ik}\left((\alpha_k + X_i\beta_k) + \sum_{l>k} y_{ik}y_{il}\psi_{kl}\right), \tag{2.5}$$

where we use that $\mu_{(0,\ldots,0)} = 0$ for identification. Clearly, the odds ratio equals μ_{y_i} as defined in (2.4) and provides the probability to observe y_i relative to the base set of choices where all choices are 0.

The association parameter ψ_{kl} is in theory an unbounded parameter and thus does not directly represent a correlation. However, log odds ratios give a direct interpretation of these association parameters. That is, it is easy to show that

$$\log\left(\frac{\Pr[Y_i = (0,\ldots,0,y_k = 1,0,\ldots,0,y_l = 1,0,\ldots,0)|X_i]\Pr[Y_i = (0,\ldots,0)|X_i]}{\Pr[Y_i = (0,\ldots,0,y_k = 1,0,\ldots,0)|X_i]\Pr[Y_i = (0,\ldots,0,y_l = 1,0,\ldots,0)|X_i]}\right) = \psi_{kl}. \tag{2.6}$$

A positive ψ_{kl} thus implies that choices k and l more often move together than apart.

The MVL model can be used to find dependencies in multivariate choices. In the next section we discuss several estimation methods to uncover these dependencies. We discuss why standard ML estimation using the joint probabilities in (2.3) is not computationally feasible in case K is large. New feasible methods are therefore introduced.

2.3 Parameter Inference

This section proposes four estimation methods for the MVL model specification defined in Section 2.2. The first approach is a standard maximum likelihood estimation procedure. This approach however is computationally infeasible when the number of choices K is large. We therefore propose three alternative novel estimation methods.

Standard ML

The first estimation method directly follows Russell and Petersen (2000). To estimate the model parameters they suggest to use the joint probabilities in (2.3). That is, Russell and Petersen (2000) use the MNL specification on the full outcome space S which results in the log-likelihood function

$$\ell^m(\theta; y) = \sum_{i=1}^{N} \log\Pr[Y_i = y_i|X_i], \tag{2.7}$$

where the joint probabilities $\Pr[Y_i = y_i|X_i]$ are given in (2.3). Further, θ summarizes all model parameters. To distinguish between the several methods we add the superscript m to the likelihood function. Standard errors of the estimator can be obtained in the same way as for standard MNL models, see, for example Amemiya (1985).

This estimation approach is very suitable when the number of choices K is small. However, the number of alternatives S increases exponentially with K. For example, 10 binary choices already lead to $2^{10} = 1024$ potential outcomes of Y_i. This leads to very small probabilities in (2.3) and a sum of many terms in the denominator, which may both lead to computational problems. Furthermore, the computation time of the probabilities and hence the log-likelihood function will increase rapidly with the number of choices. The dominating factor in the time spent computing the log likelihood function for a single observation in (2.7) is the sum over the exponents, which has order of complexity 2^K. We next propose three alternative novel estimation methods which avoid the computation of all joint probabilities.

Stratified Importance Sampling

The first alternative method reduces the number of elements in the denominator and thereby avoids large summations and the evaluation of small probabilities. To achieve this we use a stratified subset of the full outcome space for parameter estimation, where the selection probabilities for outcomes differ. Straightforwardly using such a selection may however result in an inconsistent ML estimator. We use the correction term of Ben-Akiva and Lerman (1985, Section 9.3) to correct for the stratification. This correction term is related to the sampling probability of the subset.

Formally, let D_i be a subset of the full outcome space S. We know from McFadden (1978) that maximization of the conditional log-likelihood

$$\ell^s(\theta; y) = \sum_{i=1}^{N} \log \Pr[Y_i = y_i|D_i, X_i] \tag{2.8}$$

yields consistent parameter estimates if $y_i \in D_i$. From Bayes' theorem we can write

$$\begin{aligned}
\Pr[Y_i = y_i|D_i, X_i] &= \frac{\Pr[Y_i = y_i|X_i] \Pr[D_i|Y_i = y_i, X_i]}{\sum_{d_i \in D_i} \Pr[Y_i = d_i|X_i] \Pr[D_i|Y_i = d_i, X_i]} \\
&= \frac{\exp\left(\mu_{y_i} + \log\left(\Pr[D_i|Y_i = y_i, X_i]\right)\right)}{\sum_{d_i \in D_i} \exp\left(\mu_{d_i} + \log\left(\Pr[D_i|Y_i = d_i, X_i]\right)\right)},
\end{aligned} \tag{2.9}$$

where we use that $\Pr[Y_i = y_i|X_i]$ for all y_i in S follows from (2.3). Hence, the correction term in the MNL specification for using sub-sample D_i instead of full outcome space S is $\log\left(\Pr[D|Y_i = y_i, X_i]\right)$.

To select an appropriate sub-sample D_i we follow Ben-Akiva and Lerman (1985). They propose to use stratified importance sampling [SIS] for the creation of the subset D_i and to find the values for the correction term. This selection method creates disjoint strata containing comparable alternatives. One randomly selects (with equal probabilities) a fixed number of alternatives within each stratum. For stratum r we select n_r alternatives. For the stratum that contains y_i we make sure that y_i is contained in the selected set.

Specifically, we create strata of singles, pairs, triplets et cetera in the multivariate binary choice data. Even though there may be many triplets, SIS allows us to limit the number of triplets we actually need to consider.

Formally, let R be the number of disjoint strata and let q_r be the stratum-specific probability to be in subset D_i based on the fixed amount of alternatives to be drawn. This probability equals n_r divided by the number of alternatives in stratum r. Then, referring to Ben-Akiva and Lerman (1985), $\Pr[D_i|Y_i = y_i, X_i] \propto 1/q_{r(y_i)}$, where $r(y_i)$ is the stratum containing the joint set of binary choices under consideration.

Hence, the correction term equals the negative logarithm of the stratum-specific selection probabilities. The joint probabilities in (2.9) are then given by

$$\Pr[Y_i = y_i|D_i, X_i] = \frac{\exp\left(\mu_{y_i} - \log\left(q_{r(y_i)}\right)\right)}{\sum_{d_i \in D_i} \exp\left(\mu_{d_i} - \log\left(q_{r(d_i)}\right)\right)}. \tag{2.10}$$

Replacing the joint probabilities in (2.7) by (2.10) provides a stratified log-likelihood. The stratified importance sampling ML estimator is consistent but there is loss in efficiency compared to full ML due to the sampling.

It is easy to see the advantages of this approach over the standard ML approach of Russell and Petersen (2000). Using only a subset D_i in stratified importance sampling reduces the dimension in the MVL model and thereby avoids the large summation in the denominator of (2.3). The order of complexity of a likelihood contribution calculation reduces from 2^K to the size of D_i, which can be chosen considerably smaller than 2^K. Furthermore, an optimal choice of strata R and sampling probabilities q_r will not imply large efficiency losses. Nonetheless, small sampling probabilities q_r decreases computation time but increases efficiency loss. A Monte Carlo study has to shed light on the effect of the size of D_i on efficiency losses. In the remainder of this section we introduce two alternative novel estimation methods.

Composite Conditional Likelihood

Given the structure of the multivariate logit model it is possible to use composite conditional likelihood (Lindsay, 1988) for parameter estimation. Where both the method by Russell and Petersen (2000) and the method proposed in the previous paragraph write the MVL model as a Multinomial Logit specification on a large outcome space, the CCL representation uses the conditional probabilities in (2.1) as separate, nonetheless correlated, choices. Hence, CCL avoids summation over the complete outcome space. It can be shown that the CCL approach provides consistent estimators at the cost of a loss in efficiency (Varin et al., 2011).

Following Molenberghs and Verbeke (2005, Chapter 12), the conditional probabilities in (2.1) lead to the composite log-likelihood function for the MVL model, that is,

$$\begin{aligned} \ell^c(\theta; y) &= \sum_{i=1}^{N} \ell^c(\theta; y_i) = \sum_{i=1}^{N} \sum_{k=1}^{K} \ell^c(\theta; y_{ik}) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{K} \log \Pr[Y_{ik} = y_{ik} | y_{il} \text{ for } l \neq k, X_i], \end{aligned} \tag{2.11}$$

where the superscript c stands for CCL. The estimator $\hat{\theta}$ which follows from maximizing (2.11) is consistent as $N \to \infty$ (Varin et al., 2011).

Varin et al. (2011) furthermore show that standard errors in CCL can be computed using the Godambe (1960) information matrix, which has a sandwich form and equals

$$G^c_{\hat{\theta}} = H^c_{\hat{\theta}} \left(J^c_{\hat{\theta}} \right)^{-1} H^c_{\hat{\theta}} \tag{2.12}$$

with

$$H^c_{\hat{\theta}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \nabla \ell^c(\hat{\theta}; y_{ik}) \nabla \ell^{c\prime}(\hat{\theta}; y_{ik}) \text{ and } J^c_{\hat{\theta}} = \frac{1}{N} \sum_{i=1}^{N} \nabla \ell^c(\hat{\theta}; y_i) \nabla \ell^{c\prime}(\hat{\theta}; y_i), \tag{2.13}$$

where $\nabla \ell^c(\hat{\theta}; y_{ik})$ and $\nabla \ell^c(\hat{\theta}; y_i)$ denote the first-order derivatives of the corresponding log-likelihood contributions in (2.11). The covariance matrix of the parameter estimates then follows from $(-G^c_{\hat{\theta}})^{-1}$.

Although the composite conditional likelihood does not correspond to the correct likelihood function, it still takes dependencies in the MVL model into account. The advantage over the full multinomial representation in (2.3) is that CCL avoids the large summation in the denominator. The order of complexity for a likelihood contribution is further reduced to K because of the separation of conditional choices. It is therefore possible to compute CCL even when there is a large number of choices.

Nonetheless, since the composite instead of the true likelihood function is used, the estimator is not efficient. A Monte Carlo study in Section 2.4 will however show a rather small and acceptable efficiency loss.

Generalized Method of Moments

The final estimation method we consider for the multivariate logit model is generalized method of moments (Hansen, 1982). To reduce the computation time we base the moment conditions only on the conditional probabilities. Assuming exogeneity of the explanatory variables, the moment conditions

$$\begin{aligned} \mathbb{E}(Y_{ik} - \Pr[Y_{ik} = 1 | y_{il} \text{ for } l \neq k, X_i]) &= 0 \quad \forall\, k = 1, \ldots, K, \\ \mathbb{E}((Y_{ik} - \Pr[Y_{ik} = 1 | y_{il} \text{ for } l \neq k, X_i])X_i) &= 0 \quad \forall\, k = 1, \ldots, K, \\ \mathbb{E}((Y_{ik} - \Pr[Y_{ik} = 1 | y_{il} \text{ for } l \neq k, X_i])Y_{il}) &= 0 \quad \forall\, l \neq k \end{aligned} \tag{2.14}$$

are valid to estimate the parameters in θ. We denote the sample analogue of these moment conditions for observation i by $m_i(\theta)$, which is a $(p+K) \times K$-dimensional vector.

The number of moment conditions equals $(p+K) \times K$. When $K > 1$, the number of moment conditions exceeds the number of parameters in the model and we use a two-step GMM approach (Cameron and Trivedi, 2005, Chapter 6). First, we estimate the parameters assigning equal weight to all moment conditions. In the second step, we optimally weigh the moment conditions according to the covariance matrix of the moment conditions to obtain the final parameter estimates. That is, in the second step we solve

$$\min_{\theta} M(\theta)' W M(\theta), \tag{2.15}$$

where $M(\theta) = \frac{1}{N}\sum_{i=1}^{N} m_i(\theta)$. The weighting matrix W is estimated as the matrix $(\frac{1}{N}\sum_{i=1}^{N} m_i(\theta) m_i(\theta)')^{-1}$ evaluated at the first round estimate of θ, see, for example, Cameron and Trivedi (2005, Chapter 6.3).

The covariance matrix of the parameter estimates from GMM follows from

$$\left(H_{\hat{\theta}}^{g\prime} \left(J_{\hat{\theta}}^{g} \right)^{-1} H_{\hat{\theta}}^{g} \right)^{-1} \tag{2.16}$$

with $H_{\hat{\theta}}^{g} = \sum_{i=1}^{N} \nabla m_i(\hat{\theta})$ and $J_{\hat{\theta}}^{g} = \sum_{i=1}^{N} m_i(\hat{\theta}) m_i'(\hat{\theta})$ where the superscript g stands for GMM.

The GMM approach uses conditional probabilities (2.1) instead of joint probabilities (2.3) and hence the large summation in the denominator of (2.3) is avoided. GMM therefore has the same computational advantages as the CCL approach. The order of complexity for a single observation equals the number of moment conditions. Hence, this is lower than 2^K if $K > 4$ and p reasonably small. As the suggested GMM approach has more moment conditions than parameters it is possible to use a standard test for over-identifying restrictions to test for the validity of the MVL model specification.

In sum, in this section we have proposed four parameter estimation methods for the multivariate logit model. Since the standard ML method is computationally infeasible when the number of choices is large, we have proposed three novel estimation methods. In the next section we compare these new estimation methods with the standard ML approach in a Monte Carlo study. We focus on small-sample bias, loss in efficiency and computation time for several numbers of correlated binary choices K and sample sizes N.

2.4 Monte Carlo Study

In this section we conduct a Monte Carlo study to investigate the properties of the four estimation methods described in the previous sections. First, we compare computation times of the four methods. Second, we examine small-sample bias and efficiency losses by looking at the average parameter estimates and the root mean squared error [RMSE] over the replications. Since the standard ML method uses the full information likelihood function, this method is expected to be most efficient. We compare the three alternative novel estimation methods to this method to analyze loss in efficiency. Finally, we check whether standard errors provided by the methods allow for valid inference in small samples.

For our Monte Carlo study we consider the MVL specification in (2.3) and (2.4). The number of choices is either small ($K = 4$), medium ($K = 8$) or large ($K = 12$). We consider a relatively small sample size ($N = 500$) and a large sample ($N = 5000$). As explanatory variables X_i we take two positively correlated random variables; one continuous and one discrete. Both variables are drawn from a bivariate normal distribution with variances 0.25 and correlation 0.75 and the second variable is made discrete based on a zero threshold. To avoid the need to consider many different data generating processes [DGPs], the DGP parameters are chosen in such a way that different types of correlation structures occur within our set of K binary variables, see Tables 2.2 to 2.4 for the values of the DGP-parameters. For all K, positive and negative as well as large and small association parameters are used. Note that the size of the association parameters depends on K and thus differs over K. The GMM approach uses the discussed two-step estimator. For the stratified sampling approach we have to choose R and q_r. Since the sets of binary choices within a stratum should be comparable, we create

Table 2.1: Average computation time over 100 replications (1000 observations)[a]

	Estimation method				
Number of choices K	ML	$SIS_{2^{K/2}}$	$SIS_{2^{K/3}}$	CCL	GMM
4	0.79	1.02	0.89	0.25	1.22
8	37.33	15.89	8.17	1.66	7.25
12	1538.94	200.76	70.94	5.57	33.73

[a] In seconds in Matlab R2013a on a Quad-Core Intel Xeon 2.67Ghz processor (8GB RAM) running Windows 7 64 bits

strata of singles, pairs, triplets, et cetera. An intuitive choice for q_r is the relative fraction of stratum r in the data. We consider two alternatives: one where the size of subset D_i is $2^{K/2}$ and one where it is $2^{K/3}$.

All estimation methods are implemented in Matlab R2013a on a quad-core Intel Xeon 2.67Ghz processor with 8GB RAM. Before we discuss the results of the Monte Carlo study, we first focus on computation time. Table 2.1 displays the average computation time over 100 replications and $N = 1000$ observations for different values of K, where we use the DGPs from Tables 2.2 to 2.4. Since large summations in the denominator of (2.3) and small joint probabilities do not occur for small K, standard ML estimation is still computationally feasible. However, for larger K, differences in computation time grow rapidly. For instance, the computation time for standard ML when $K = 12$ is on average 25.6 minutes and the other three methods have a clear advantage. The computation time of CCL is more than 275 times faster (only 5.6 seconds). These computation times are in line with the (objective) order of complexity presented in Section 2.3. If the small-sample bias and losses in efficiency are both small, the alternative estimation methods are sound alternatives for parameter estimation in the large MNL specification with large K. Note that the difference in computation time will further increase if we include more explanatory variables in the model or consider even larger K.

Tables 2.2 to 2.4 display the average and RMSE of the estimators over 5000 replications. Since results are highly comparable and to save space, a diverse selection of parameters from the DGPs is displayed.[1] The DGP with $N = 5000$ shows that the bias is quite small for all estimation methods. For small sample sizes, the deviation of the parameter estimates from the DGP values is larger. Nonetheless, all methods find comparably accurate estimates. Our newly introduced estimation methods thus are capable of finding estimates comparable to the regular likelihood approach.

To further analyze the loss in efficiency between the three novel estimation methods and standard ML, we consider best and worst cases of the RMSEs across all parameters, see Tables 2.2 to 2.4. As expected, standard ML is most efficient. The subset approach used in SIS causes a loss of information

[1] The results for the other parameters are similar and available upon request.

and thereby an increase in RMSE. Obviously, the smaller the subset, the larger the loss in efficiency. In the best and worst case, the RMSE of ML and SIS with a subset D of size $2^{K/2}$ differ 3.7 and 7.0 percent, respectively. The smaller subset of size $2^{K/3}$ yields efficiency losses between 12.0 and 20.4 percent. For CCL and GMM, only small efficiency losses occur. The differences of GMM with ML in terms of RMSE are between 0.02 and 7.3 percent. These differences are smallest for the parameters of the covariates. For CCL, the minimum and maximum differences are only 0.1 and 0.9 percent, respectively.

Table 2.2: Average parameter estimates and RMSE in a simulation study with 4 binary choices (5000 replications)[a]

	DGP	ML		$SIS_{2^{K/2}}$		$SIS_{2^{K/3}}$		CCL		GMM	
$N = 500$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	-0.35	-0.358	0.230	-0.354	0.257	-0.365	0.298	-0.358	0.230	-0.381	0.239
β_2	-1	-1.018	0.277	-1.027	0.320	-1.037	0.364	-1.018	0.277	-0.990	0.274
	-0.5	-0.503	0.251	-0.508	0.286	-0.508	0.315	-0.504	0.252	-0.498	0.252
$\psi_{1,4}$	0.35	0.354	0.220	0.357	0.259	0.361	0.277	0.354	0.220	0.355	0.236
$\psi_{2,4}$	-0.9	-0.912	0.231	-0.926	0.260	-0.930	0.277	-0.913	0.231	-0.851	0.239
$\psi_{3,4}$	0.55	0.559	0.212	0.562	0.248	0.567	0.279	0.559	0.212	0.562	0.230
$N = 5000$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	-0.35	-0.350	0.071	-0.349	0.079	-0.351	0.091	-0.350	0.071	-0.353	0.071
β_2	-1	-1.003	0.085	-1.003	0.098	-1.003	0.108	-1.003	0.086	-0.998	0.085
	-0.5	-0.499	0.077	-0.500	0.088	-0.501	0.095	-0.499	0.077	-0.499	0.076
$\psi_{1,4}$	0.35	0.351	0.068	0.352	0.079	0.353	0.085	0.351	0.068	0.352	0.069
$\psi_{2,4}$	-0.9	-0.902	0.071	-0.904	0.081	-0.903	0.084	-0.902	0.071	-0.894	0.070
$\psi_{3,4}$	0.55	0.551	0.067	0.552	0.078	0.553	0.086	0.551	0.067	0.551	0.069

[a] To save space we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

Table 2.3: Average parameter estimates and RMSE in a simulation study with 8 binary choices (5000 replications)[a]

	DGP	ML		$SIS_{2K/2}$		$SIS_{2K/3}$		CCL		GMM	
$N = 500$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	-0.95	-0.972	0.269	-0.974	0.286	-0.973	0.316	-0.972	0.270	-1.014	0.287
β_3	-1	-1.024	0.330	-1.032	0.352	-1.050	0.393	-1.026	0.333	-0.986	0.331
	-0.5	-0.511	0.295	-0.517	0.310	-0.521	0.345	-0.512	0.296	-0.504	0.299
$\psi_{1,8}$	0	-0.009	0.262	-0.008	0.275	-0.011	0.299	-0.009	0.263	0.003	0.271
$\psi_{2,7}$	0.15	0.146	0.257	0.148	0.269	0.151	0.294	0.146	0.257	0.152	0.266
$\psi_{3,5}$	-0.9	-0.928	0.296	-0.936	0.309	-0.959	0.331	-0.931	0.297	-0.824	0.302
$N = 5000$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	-0.95	-0.949	0.082	-0.949	0.087	-0.949	0.096	-0.949	0.082	-0.954	0.084
β_3	-1	-1.003	0.099	-1.004	0.105	-1.005	0.115	-1.003	0.099	-0.994	0.100
	-0.5	-0.501	0.090	-0.502	0.093	-0.503	0.103	-0.501	0.090	-0.499	0.090
$\psi_{1,8}$	0	-0.001	0.080	-0.001	0.084	-0.001	0.090	-0.001	0.080	0.002	0.082
$\psi_{2,7}$	0.15	0.149	0.079	0.149	0.082	0.148	0.087	0.149	0.079	0.150	0.080
$\psi_{3,5}$	-0.9	-0.905	0.092	-0.906	0.094	-0.908	0.101	-0.905	0.092	-0.875	0.097

[a] To save space we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

Table 2.4: Average parameter estimates and RMSE in a simulation study with 12 binary choices (5000 replications)[a]

	DGP	ML[b]		$SIS_{2K/2}$[b]		$SIS_{2K/3}$		CCL		GMM	
$N = 500$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	-1.55	–	–	–	–	-1.602	0.368	-1.591	0.314	-1.645	0.347
β_4	-1	–	–	–	–	-1.074	0.451	-1.040	0.386	-0.995	0.390
	-0.5	–	–	–	–	-0.525	0.401	-0.508	0.340	-0.518	0.352
$\psi_{3,12}$	-0.35	–	–	–	–	-0.405	0.432	-0.390	0.397	-0.346	0.395
$\psi_{5,10}$	0.15	–	–	–	–	0.136	0.398	0.133	0.368	0.114	0.371
$\psi_{7,8}$	0.55	–	–	–	–	0.570	0.390	0.554	0.349	0.486	0.374
$N = 5000$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	-1.55	–	–	–	–	-1.558	0.106	-1.555	0.094	-1.561	0.097
β_4	-1	–	–	–	–	-1.007	0.128	-1.005	0.116	-0.993	0.115
	-0.5	–	–	–	–	-0.503	0.117	-0.502	0.103	-0.505	0.103
$\psi_{1,4}$	-0.35	–	–	–	–	-0.355	0.121	-0.352	0.116	-0.341	0.116
$\psi_{2,4}$	0.15	–	–	–	–	0.151	0.113	0.150	0.107	0.139	0.109
$\psi_{3,4}$	0.55	–	–	–	–	0.548	0.111	0.547	0.103	0.519	0.110

[a] To save space we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

[b] As estimation for ML and $SIS_{2K/2}$ take too long (see Table 2.1) we do not include them in the 5000 replications simulation.

In practice one usually opts for the most efficient approach. However, the estimation method should also be computationally feasible such that parameter estimates can be obtained in a reasonable amount of time. The large summation over all possible alternatives in the standard ML method may lead to numerical problems and long computation times for large K. CCL and GMM seem to be useful alternatives for standard ML and produce useful parameter estimates in little time. The small-sample bias is similar and the loss in efficiency is rather small. For SIS, there is a clear trade-off between the size of the subset and the loss in efficiency.

Table 2.5: Empirical size of the distribution of the four estimators of the MVL model with 4 binary choices (5000 observations, 5000 replications)[a]

		Percentiles					
Model	Theoretical	0.025	0.05	0.1	0.9	0.95	0.975
ML	α_1	0.026	0.052	0.099	0.896	0.949	0.977
	β_2	0.025	0.048	0.098	0.894	0.947	0.972
		0.024	0.048	0.097	0.902	0.950	0.975
	$\psi_{1,4}$	0.024	0.050	0.099	0.901	0.949	0.976
	$\psi_{2,4}$	0.023	0.047	0.097	0.896	0.946	0.972
	$\psi_{3,4}$	0.026	0.052	0.099	0.898	0.949	0.977
$SIS_{2^{K/2}}$	α_1	0.028	0.051	0.100	0.897	0.949	0.975
	β_2	0.024	0.049	0.096	0.898	0.947	0.972
		0.024	0.049	0.098	0.898	0.949	0.975
	$\psi_{1,4}$	0.027	0.051	0.103	0.900	0.953	0.975
	$\psi_{2,4}$	0.023	0.046	0.096	0.892	0.944	0.972
	$\psi_{3,4}$	0.025	0.050	0.100	0.900	0.949	0.976
$SIS_{2^{K/3}}$	α_1	0.026	0.051	0.098	0.896	0.948	0.974
	β_2	0.022	0.049	0.099	0.899	0.948	0.975
		0.025	0.047	0.096	0.906	0.952	0.977
	$\psi_{1,4}$	0.024	0.049	0.097	0.899	0.949	0.975
	$\psi_{2,4}$	0.025	0.050	0.101	0.898	0.948	0.973
	$\psi_{3,4}$	0.027	0.049	0.101	0.895	0.946	0.975
CCL	α_1	0.027	0.052	0.099	0.896	0.948	0.977
	β_2	0.025	0.049	0.098	0.893	0.946	0.972
		0.025	0.048	0.098	0.903	0.950	0.975
	$\psi_{1,4}$	0.025	0.050	0.099	0.900	0.949	0.974
	$\psi_{2,4}$	0.023	0.048	0.099	0.895	0.945	0.972
	$\psi_{3,4}$	0.025	0.053	0.099	0.898	0.949	0.977
GMM	α_1	0.029	0.057	0.106	0.888	0.943	0.972
	β_2	0.027	0.053	0.105	0.889	0.942	0.970
		0.027	0.050	0.100	0.903	0.950	0.973
	$\psi_{1,4}$	0.032	0.062	0.111	0.888	0.940	0.969
	$\psi_{2,4}$	0.033	0.063	0.116	0.881	0.933	0.965
	$\psi_{3,4}$	0.032	0.061	0.111	0.885	0.940	0.970

[a] To save space we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

Table 2.6: Empirical size of the distribution of the four estimators of the MVL model with 8 binary choices (5000 observations, 5000 replications)[a]

		Percentiles					
Model	Theoretical	0.025	0.05	0.1	0.9	0.95	0.975
ML	α_1	0.022	0.048	0.098	0.900	0.948	0.972
	β_3	0.021	0.044	0.099	0.899	0.949	0.978
		0.026	0.051	0.101	0.899	0.954	0.977
	$\psi_{1,8}$	0.025	0.048	0.096	0.901	0.952	0.976
	$\psi_{2,7}$	0.025	0.052	0.104	0.891	0.947	0.975
	$\psi_{3,5}$	0.022	0.048	0.100	0.898	0.944	0.974
$SIS_{2^{K/2}}$	α_1	0.027	0.052	0.102	0.900	0.949	0.975
	β_3	0.023	0.047	0.099	0.900	0.948	0.976
		0.026	0.050	0.102	0.892	0.950	0.976
	$\psi_{1,8}$	0.027	0.053	0.096	0.899	0.952	0.978
	$\psi_{2,7}$	0.025	0.056	0.103	0.894	0.948	0.975
	$\psi_{3,5}$	0.025	0.047	0.093	0.893	0.945	0.974
$SIS_{2^{K/3}}$	α_1	0.023	0.050	0.105	0.902	0.948	0.976
	β_3	0.022	0.045	0.098	0.897	0.948	0.973
		0.027	0.050	0.100	0.900	0.954	0.979
	$\psi_{1,8}$	0.026	0.047	0.098	0.899	0.951	0.977
	$\psi_{2,7}$	0.023	0.049	0.099	0.898	0.947	0.975
	$\psi_{3,5}$	0.025	0.049	0.098	0.890	0.946	0.974
CCL	α_1	0.023	0.048	0.100	0.900	0.948	0.972
	β_3	0.022	0.044	0.100	0.898	0.949	0.976
		0.026	0.051	0.103	0.899	0.952	0.977
	$\psi_{1,8}$	0.026	0.049	0.099	0.896	0.951	0.974
	$\psi_{2,7}$	0.027	0.054	0.105	0.888	0.945	0.975
	$\psi_{3,5}$	0.024	0.049	0.100	0.897	0.942	0.970
GMM	α_1	0.029	0.057	0.109	0.887	0.941	0.967
	β_3	0.028	0.054	0.107	0.886	0.941	0.970
		0.029	0.055	0.105	0.892	0.949	0.976
	$\psi_{1,8}$	0.035	0.060	0.117	0.874	0.931	0.961
	$\psi_{2,7}$	0.039	0.069	0.119	0.868	0.931	0.963
	$\psi_{3,5}$	0.034	0.064	0.121	0.873	0.930	0.958

[a] To save space we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

Table 2.7: Empirical size of the distribution of the four estimators of the MVL model with 12 binary choices (5000 observations, 5000 replications)[a]

		Percentiles					
Model	Theoretical	0.025	0.05	0.1	0.9	0.95	0.975
$SIS_{2K/3}$	α_1	0.025	0.048	0.093	0.900	0.950	0.975
	β_4	0.023	0.051	0.098	0.903	0.957	0.977
		0.023	0.044	0.095	0.898	0.949	0.975
	$\psi_{3,12}$	0.024	0.046	0.093	0.902	0.949	0.975
	$\psi_{5,10}$	0.021	0.046	0.094	0.901	0.953	0.977
	$\psi_{7,8}$	0.024	0.042	0.094	0.904	0.947	0.974
CCL	α_1	0.025	0.050	0.095	0.894	0.948	0.974
	β_4	0.024	0.051	0.106	0.894	0.946	0.975
		0.024	0.048	0.097	0.902	0.949	0.971
	$\psi_{3,12}$	0.024	0.048	0.098	0.891	0.947	0.974
	$\psi_{5,10}$	0.023	0.049	0.101	0.895	0.948	0.974
	$\psi_{7,8}$	0.025	0.050	0.098	0.898	0.950	0.972
GMM	α_1	0.036	0.066	0.119	0.876	0.935	0.965
	β_4	0.030	0.065	0.120	0.882	0.938	0.967
		0.028	0.055	0.102	0.892	0.943	0.968
	$\psi_{3,12}$	0.044	0.076	0.127	0.862	0.918	0.953
	$\psi_{5,10}$	0.043	0.069	0.129	0.862	0.920	0.954
	$\psi_{7,8}$	0.045	0.072	0.125	0.870	0.925	0.954

[a] To save space we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

Apart from bias and efficiency, we also consider the validity of the standard errors with respect to significance testing of the model parameters. Tables 2.5 to 2.7 display the empirical size of the t-test for $N = 5000$ for both tails of the t-statistic. The table shows that size distortions are rather small. The largest size distortions are found for the GMM approach. For example, a theoretical 90 percent confidence interval for $\psi_{3,12}$ in GMM turns out to have a coverage of 84.2%. This size distortion is still acceptable. For the other approaches the size distortions are smaller. The same coverage probability is 89.9% for the CCL approach. Unreported results show that even for small N size distortions of ML, SIS and CCL are still negligible. Hence, hypothesis tests can be carried out in the usual manner for these estimation methods. In accordance with existing literature (Altonji and Segal, 1996), size distortion for the GMM approach are larger in small samples.

In sum, the Monte Carlo study shows that the novel estimation methods are sound alternatives for the regular likelihood approach. Where computation times in standard ML increase exponentially over the number of choices, the computation time stays limited using CCL, GMM or SIS. Further, small-sample biases are comparable to full ML and efficiency losses are rather small and acceptable. Given the win in computation time, the avoidance of numerical problems, small small-sample biases and negligible losses in efficiency, CCL is the most promising alternative estimation method.

2.5 Application

In this section we illustrate the use of an MVL model with many choices. We consider survey data of 2046 individuals on store visits in a particular Dutch specialized shopping mall. Visits to different stores are likely to be correlated and hence, it is convenient to model these simultaneous decisions using a multivariate logit specification. In this application we consider simultaneous choices for ten different stores. All stores fall under the general theme of home decoration and do-it-yourself. Table 2.8 details the types of stores. Our dependent variable can take $2^{10} = 1024$ different values. As explanatory variables we have *Family size*, *Age*, *Gender*, *Income*, *Number of visits* and *Appreciation* of the shopping mall.

The simulation study in Section 2.4 showed that for this size of the outcome space, large differences in computation time occur. Hence, one may not be willing to use standard Maximum Likelihood estimation. Based on the simulation results we consider the CCL approach (fast and accurate) to estimate the model parameters.[2] As benchmark we will also consider the standard ML approach. The standard ML approach takes about 1.6 hours on a dual-core Intel 3.4Ghz processor with 4GB RAM which shows that this method is not very convenient if you want to investigate several model specifications. The CCL approach on the other hand only takes 2.3 minutes.

First, we test for independence among the choices for store visits. The Likelihood Ratio [LR] statistic in the maximum likelihood approach for the restriction that all $\psi = 0$ is 1373.4 (45 degrees of freedom). This statistic clearly shows that independence is rejected. Hence, we find evidence for correlations between visiting the different store types and the MVL model from Section 2.2 thus is applicable to the data. An adjusted LR-test for CCL (Varin et al., 2011) yields the same conclusion.

Tables 2.8 to 2.11 display the parameter estimates and standard errors for the two estimation methods. The parameter estimates are very similar and both methods find the same parameter estimates to be significantly different from 0. The standard errors in the CCL approach are slightly smaller than in the standard ML estimation approach but this may be due to the relatively small sample size. Unreported results show that the GMM and SIS approach also provide similar results. The results of SIS indicate that subset D_i should be large to get results close to standard ML.

The negative estimates of the choice-specific intercepts in Tables 2.8 and 2.10 show that most stores are visited only by a minority of the individuals. The order of the intercepts shows that stores selling kitchens are visited least, where stores selling building materials are visited by the most individuals.

[2] The results of the other two approaches are available upon request.

Several relations between the explanatory variables and store visits are found. For example, the more frequent visitors of the mall visit more stores selling paint/wallpaper, building materials and hardware. These can be seen as the fanatic handymen. Furthermore, visitors who very much appreciate the mall are more likely to also buy their furniture, lamps and floor and wall decorations at this shopping mall.

The association parameters in Table 2.11 show the relations between the visits to different stores. Clear interpretations can be given. For example, individuals who visit a store selling an odd jobs article (paint/wallpaper, building materials or hardware) are likely also to visit other odd jobs stores. The same holds for stores selling lamps, curtains/carpets and furniture since the corresponding association parameters are positive. Negative and significant association parameters are for instance found for the combination hardware and curtains/carpets. Apparently, individuals seem to be unlikely to visit both these store types in this shopping mall.

In sum, the MVL model gives understandable and interpretable parameter estimates for the data of store visits in a Dutch shopping mall. Furthermore, the standard ML and CCL approach yield very similar estimation results and conclusions. The clear advantage of the CCL approach is the time it takes to obtain parameter estimates with small loss in efficiency. The reduction in computation time is large, and with the CCL method it becomes feasible to easily consider several model specifications. In case the number of stores would have been larger, ML estimation would break down while CCL could still be used.

Table 2.8: Parameter estimates and standard errors of the MVL model using the standard ML method for shopping mall data

	Bed		Paint/Wallpaper		Materials		Hardware		Furniture		Lamps		Garden		Curtain/Carpet		Kitchen		Other	
	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
intercept	-1.373	0.241	-1.649	0.245	-0.699	0.253	-1.339	0.256	-1.854	0.266	-2.196	0.248	-2.547	0.288	-1.459	0.243	-3.042	0.350	-2.199	0.437
family size	0.163	0.043	-0.060	0.043	0.059	0.047	-0.063	0.047	0.028	0.045	0.040	0.042	-0.019	0.048	-0.012	0.043	-0.160	0.061	0.078	0.075
young (< 35)	-0.064	0.118	0.101	0.121	-0.035	0.130	-0.165	0.129	0.213	0.125	0.237	0.118	0.143	0.130	-0.099	0.121	-0.061	0.165	0.442	0.214
old (> 54)	-0.362	0.166	-0.095	0.156	-0.222	0.170	-0.039	0.175	-0.323	0.186	-0.185	0.163	0.280	0.173	0.041	0.162	-0.190	0.234	-0.007	0.332
female	0.086	0.106	0.128	0.105	-0.086	0.117	-0.136	0.114	-0.193	0.116	-0.126	0.106	0.049	0.118	0.165	0.107	-0.001	0.150	0.121	0.214
low income	0.074	0.125	0.094	0.123	-0.100	0.133	-0.097	0.129	-0.025	0.135	0.143	0.124	0.034	0.136	0.098	0.127	-0.228	0.181	0.167	0.240
high income	-0.219	0.170	-0.144	0.173	0.021	0.188	0.598	0.188	0.029	0.183	-0.095	0.171	-0.225	0.194	-0.448	0.175	-0.293	0.250	0.401	0.295
# visits	0.003	0.003	0.023	0.003	0.018	0.004	0.030	0.004	-0.001	0.003	0.007	0.003	0.005	0.003	0.002	0.003	0.004	0.003	0.009	0.006
positive apprec.	-0.113	0.153	0.098	0.152	0.239	0.165	0.066	0.162	0.495	0.177	0.808	0.163	0.359	0.190	0.363	0.154	-0.102	0.234	0.292	0.306
negative apprec.	-0.142	0.216	-0.123	0.214	0.031	0.228	-0.065	0.225	-0.109	0.263	0.231	0.233	0.152	0.261	-0.106	0.223	0.000	0.330	0.281	0.386

Table 2.9: Estimates of the association parameters and standard errors of the MVL model using the standard ML method for shopping mall data

	Bed		Paint/Wallpaper		Materials		Hardware		Furniture		Lamps		Garden		Curtain/Carpet		Kitchen	
	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
Paint/Wallpaper	0.101	0.111																
Materials	-0.367	0.121	0.584	0.117														
Hardware	-0.162	0.121	0.484	0.115	1.739	0.116												
Furniture	0.656	0.115	-0.268	0.126	-0.413	0.130	-0.109	0.131										
Lamps	0.505	0.108	0.635	0.108	-0.357	0.119	0.366	0.118	0.435	0.120								
Garden	-0.070	0.127	0.274	0.123	0.341	0.141	0.529	0.134	0.075	0.137	0.708	0.123						
Curtain/Carpet	0.783	0.106	0.794	0.111	-0.146	0.120	-0.377	0.120	0.782	0.118	0.468	0.108	-0.183	0.130				
Kitchen	0.451	0.163	0.201	0.169	0.873	0.201	0.406	0.184	0.876	0.169	0.139	0.165	0.538	0.154	0.171	0.173		
Other	-0.200	0.235	-0.704	0.278	-0.890	0.264	-0.611	0.289	-0.466	0.271	-0.333	0.251	-0.128	0.277	-0.449	0.249	0.423	0.345

Table 2.10: Parameter estimates and standard errors of the MVL model using the CCL method for shopping mall data

	Bed		Paint/Wallpaper		Materials		Hardware		Furniture		Lamps		Garden		Curtain/Carpet		Kitchen		Other	
	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
intercept	-1.373	0.234	-1.647	0.237	-0.697	0.244	-1.337	0.248	-1.854	0.258	-2.195	0.241	-2.547	0.281	-1.459	0.237	-3.042	0.351	-2.199	0.434
family size	0.162	0.041	-0.055	0.042	0.066	0.045	-0.055	0.045	0.027	0.043	0.041	0.041	-0.018	0.046	-0.013	0.041	-0.161	0.060	0.078	0.074
young (<35)	-0.065	0.114	0.102	0.117	-0.034	0.128	-0.164	0.125	0.213	0.121	0.237	0.114	0.143	0.126	-0.099	0.117	-0.061	0.162	0.442	0.209
old (>54)	-0.362	0.160	-0.094	0.150	-0.222	0.164	-0.039	0.166	-0.323	0.180	-0.185	0.158	0.280	0.166	0.041	0.155	-0.190	0.228	-0.007	0.308
female	0.086	0.103	0.129	0.102	-0.085	0.113	-0.135	0.110	-0.193	0.112	-0.125	0.103	0.049	0.114	0.165	0.104	-0.001	0.145	0.121	0.208
low income	0.074	0.120	0.094	0.119	-0.100	0.129	-0.097	0.126	-0.025	0.130	0.143	0.119	0.034	0.131	0.098	0.121	-0.228	0.176	0.167	0.229
high income	-0.219	0.163	-0.144	0.167	0.021	0.183	0.599	0.180	0.029	0.177	-0.095	0.165	-0.225	0.188	-0.448	0.168	-0.293	0.242	0.401	0.288
# visits	0.003	0.003	0.021	0.003	0.015	0.003	0.026	0.004	-0.001	0.003	0.007	0.003	0.005	0.002	0.002	0.002	0.004	0.003	0.008	0.006
positive apprec.	-0.113	0.150	0.099	0.147	0.241	0.159	0.068	0.158	0.495	0.169	0.808	0.158	0.359	0.183	0.363	0.150	-0.103	0.224	0.292	0.303
negative apprec.	-0.142	0.210	-0.122	0.208	0.031	0.220	-0.064	0.218	-0.110	0.254	0.231	0.226	0.152	0.253	-0.106	0.218	0.000	0.310	0.281	0.381

Table 2.11: Estimates of the association parameters and standard errors of the MVL model using the CCL method for shopping mall data

	Bed		Paint/Wallpaper		Materials		Hardware		Furniture		Lamps		Garden		Curtain/Carpet		Kitchen	
	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
Paint/Wallpaper	0.102	0.107																
Materials	-0.366	0.114	0.586	0.114														
Hardware	-0.162	0.114	0.486	0.111	1.741	0.110												
Furniture	0.656	0.110	-0.268	0.121	-0.413	0.125	-0.109	0.126										
Lamps	0.505	0.105	0.636	0.105	-0.356	0.115	0.367	0.114	0.435	0.115								
Garden	-0.070	0.121	0.275	0.119	0.342	0.134	0.529	0.129	0.075	0.131	0.708	0.118						
Curtain/Carpet	0.782	0.103	0.795	0.107	-0.145	0.116	-0.376	0.116	0.781	0.113	0.468	0.105	-0.184	0.123				
Kitchen	0.450	0.152	0.200	0.158	0.872	0.189	0.405	0.171	0.875	0.155	0.138	0.157	0.537	0.149	0.170	0.158		
Other	-0.201	0.209	-0.704	0.244	-0.890	0.230	-0.611	0.252	-0.466	0.243	-0.333	0.230	-0.128	0.257	-0.450	0.227	0.423	0.305

2.6 Conclusion

The multivariate logit model is used to model correlated simultaneous binary choices. In this chapter we proposed three novel estimation methods for this model: estimation by (i) stratified importance sampling; (ii) composite conditional likelihood; and by (iii) generalized method of moments. The new estimation methods are especially of interest when the dimension of the choice problem is large. Methods available in the literature go together with a large computational burden. The new methods in this chapter circumvent this problem.

Results from a Monte Carlo study show that the new estimation methods yield comparable small-sample biases as the standard (full information) maximum likelihood approach as proposed by Russell and Petersen (2000). Furthermore, efficiency losses compared to the full likelihood approach are rather small. Because of these findings, the decrease in computation time and avoidance of numerical problems are clear advantages of our proposed estimation methods. The composite conditional likelihood approach turns out to have the largest decrease in computation time, leads to a very small loss in efficiency, and provides accurate standard errors.

In an application, we applied the methods to store visits in a shopping mall. Multivariate binary choice data occur widely in practice. Hence, other applications in different fields of research can be given. Since the dimension of the choice problem will often be large, our methods are highly useful in applied research.

Several extensions to the current research are possible. For instance, a conditional logit specification can easily be derived. Furthermore, the association parameters can also depend on exogenous variables or be individual-specific (in panel data models). Finally, instead of binary choices, this model can be extended to a multivariate multinomial specification. The feasible estimation methods proposed in this chapter can be used in all these cases. Especially CCL is applicable to extensions of the current model specification if a clear composition of the conditional probabilities can be given.

Chapter 3

A Multivariate Model for Multinomial Choices

3.1 Introduction

It is common practice in applied research to use multinomial choice models to describe categorical data (McFadden, 1983, Chapter 24). These models are suited to describe single multinomial choices. In practice we are often dealing with multiple correlated multinomial decisions. Answers to survey questions with two or more choice possibilities are likely to be correlated. The choice for job location may be correlated with residence choice. In marketing one may be interested in dependencies of simultaneous brand choices for several product categories. Hence, simultaneous multinomial decisions occur in different areas of research.

In this chapter we propose a straightforward model to describe simultaneous multinomial decisions with potentially high dimensions. Extensions of familiar univariate models are not always suitable. That is, a correlated multinomial probit (MNP, Hausman and Wise, 1978) approach would be an obvious choice, Zhang et al. (see, for example, 2008). Nonetheless, parameter estimation of such models implies solving high-dimensional integrals. Given the computational burden in univariate MNP models (Geweke et al., 1997), frequentist inference in the multivariate counterpart is unlikely to be feasible. Another option is to use mixed logit models (Hensher and Greene, 2003) and let unobserved heterogeneity capture correlation among decisions. Again, computation implies solving integrals which becomes cumbersome when the number of simultaneous decisions increases. A nested logit specification (Maddala, 1983, Chapter 3) is perhaps a more feasible approach. However, this model is designed to describe a single choice from a large number of alternatives. The nested logit model is therefore not suitable for simultaneous decisions. Finally, one may consider a multinomial choice model for all possible combinations of choices. The number of choice combinations however easily becomes large, see also Amemiya (1978) and Ben-Akiva and Lerman (1985, Chapter 10). Clearly, model interpretation becomes difficult. Furthermore, parameter estimation becomes infeasible as probabilities get numerically small and evaluation of the likelihood requires summation over all potential outcomes.

In this chapter we want to model dependencies in a large number of simultaneous discrete decisions without the numerical problems discussed above. To fill the gap in the literature, we propose a general and novel multivariate multinomial logit [MV-MNL] specification. In essence, we extend the multivariate (binary) logit [MVL] model of Cox (1972) and Russell and Petersen (2000) to multinomial decisions. The advantages of this MV-MNL specification are that (i) the number of parameters stays limited; (ii) there is a clear interpretation of the model parameters in terms of odds ratios and that; (iii) zero restrictions on a subset of parameters result in independence between the multinomial choices.

The contribution of this chapter is threefold. First, the novel model specification is easily scalable to higher dimensions due to its special structure. Second, a quick and reliable estimation method is available which makes applying the model much more practical. That is, we can avoid the summation over all potential combinations of the multivariate multinomial choices. We consider conditional instead of joint probabilities using a composite likelihood function (Lindsay, 1988), see Chapter 2 and Bel et al. (2014) for a similar approach in MVL models. Finally, the MV-MNL specification can easily be extended to a fixed-effects specification for panel data. Parameter estimation stays feasible by using sufficient statistics in combination with the composite likelihood approach.

The model is related to the multivariate MNL specification of Amemiya (1978) and Ben-Akiva and Lerman (1985, Chapter 10) but in contrast to these specifications we explicitly focus on the dependence structure in the multinomial choices. Agresti (2007, Chapter 9) also considers dependencies in multivariate choices but is not in particular interested in the dependence structure as he proposes the evaluation of contingency tables, where individual characteristics are discarded. Burda et al. (2008) focus on numerous choices on the same attribute, but we deal with separated, nonetheless correlated, multinomial choices.

The remainder of this chapter is organized as follows. In Section 3.2, we introduce the new MV-MNL specification. We also discuss parameter identification, interpretation and parameter inference. A small Monte Carlo study shows the accuracy of the parameter estimates and a small loss in efficiency due to the use of the composite instead of the true likelihood. An extension to panel data is discussed in Section 3.3. Section 3.4 provides two illustrations of the use of MV-MNL models. The first illustration concerns a cross-sectional survey on satisfaction about life and the second illustration deals with the choice for tuna using a household panel scanner data set. Finally, Section 3.5 concludes.

3.2 Model Specification

In this section we discuss the model specification for the multivariate multinomial logit model. This model is an extension of the multivariate logit model introduced by Cox (1972) and Russell and Petersen (2000). We discuss model specification, parameter identification and interpretation of the model parameters. Section 3.2.1 shows the model representation for the choice probabilities in a simple bivariate trinomial logit model to clarify the structure and identification of parameters of the model.

Let Y_i denote the K-dimensional random variable describing the joint set of choices for individual $i = 1, \ldots, N$, defined as

$$Y_i \quad = \quad \{Y_{i1}, \ldots, Y_{iK}\}, \tag{3.1}$$

where Y_{ik} describes the k-th multinomial choice for individual i for $k = 1, \ldots, K$. $Y_{ik} = j$ if individual i chooses $j = 1, \ldots, J_k$ for the k-th choice, where J_k is the number of potential outcomes. Note that there are $\prod_{k=1}^{K} J_k$ possible realizations of the random variable Y_i. The set of possible realizations is called S.

The K choices in Y_i may be correlated. The starting point for modeling these dependencies is the conditional probabilities for each choice decision k given all choice decisions $l \neq k$, see Russell and Petersen (2000) for a binomial equivalent. These conditional probabilities are a multinomial logit function of the individual characteristics X_i, the model parameters α, β and ψ and the other choices y_{il}, that is

$$\Pr[Y_{ik} = j | y_{il} \text{ for } l \neq k, X_i] \quad = \quad \frac{\exp(Z_{ik,j})}{\sum_{l=1}^{J_k} \exp(Z_{ik,l})} \tag{3.2}$$

with

$$Z_{ik,j} \quad = \quad \alpha_{k,j} + X_i \beta_{k,j} + \sum_{l \neq k} \psi_{kl,jy_{il}}, \tag{3.3}$$

where $\alpha_{k,j}$ are alternative- and choice-specific intercepts, X_i a vector of explanatory variables with corresponding parameter vector $\beta_{k,j}$ and where $\psi_{kl,jh}$ are association parameters between choosing j

for the k-th choice and choosing h for the l-th choice. Hence, the correlation between choices y_{ik} and y_{il} is captured by the association parameter. Association means the relative change in the exponent $Z_{ik,j}$ if choices k and l move together compared to being opposite. When $\psi_{kl,jh} > 0$ this implies positive association and when $\psi_{kl,jh} < 0$ we have negative association. For $\psi_{kl,jh} = 0$ we have independence between choices y_{ik} and y_{il}.

The theorem of Besag (1974) states that all properties of the joint distribution follow from the full set of conditional distributions. We use this result to show that the conditional distributions in (3.2) imply the following multinomial logit model for the joint distribution of Y_i (see Appendix 3.A.1 for the formal proof):

$$\Pr[Y_i = y_i|X_i] = \frac{\exp(\mu_{y_i})}{\sum_{s_i \in S} \exp(\mu_{s_i})}, \tag{3.4}$$

where y_i is a possible realization from the outcome space S, and where μ_{y_i} is defined as

$$\mu_{y_i} = \sum_{k=1}^{K} \left(\alpha_{k,y_{ik}} + X_i \beta_{k,y_{ik}} + \sum_{l>k} \psi_{kl,y_{ik}y_{il}} \right). \tag{3.5}$$

It is easy to see that the equation contains α_k and β_k corresponding to the specific choice option for the k-th choice and $\psi_{kl,jh}$ corresponding to the observed choice pairs y_{ik} and y_{il}.

The role of the intercept parameters and X_i follows from the log odds ratio

$$\ln\left(\frac{\Pr[Y_i = y_i|X_i]}{\Pr[Y_i = (1,\ldots,1)'|X_i]}\right) = \sum_{k=1}^{K} \left(\alpha_{k,y_k} + X_i \beta_{k,y_k} + \sum_{l>k} \psi_{kl,y_k y_l} \right), \tag{3.6}$$

where we use that under the identification restrictions $\Pr[Y_i = (1,\ldots,1)|X_i] \propto 1$. Clearly, this odds ratio equals μ_{y_i} in (3.5) and provides the probability to observe y_i relative to the base set of choice decisions.

The parameters $\psi_{kl,jh}$ indicate the associations between choices k and l. $\psi_{kl,jh}$ is in theory an unbounded parameter and thus does not directly resemble correlation between choices j and h. To give a direct interpretation to these associations, we use log odds ratios. It is easy to show that

$$\psi_{kl,jh} = \ln\left(\frac{\Pr[Y_i = (1,\ldots,1,y_{ik}=j,1,\ldots,1,y_{il}=h,1,\ldots,1)'|X_i]\,\Pr[Y_i = (1,\ldots,1)'|X_i]}{\Pr[Y_i = (1,\ldots,1,y_{ik}=j,1,\ldots,1)'|X_i]\,\Pr[Y_i = (1,\ldots,1,y_{il}=h,1,\ldots,1)'|X_i]}\right). \tag{3.7}$$

Hence, a positive $\psi_{kl,jh}$ implies that the choices j and h more often move together than apart. Hence, this indeed implies positive $\psi_{kl,jh}$ for positive correlations and negative association parameters for negative correlations.

Various restrictions are needed for identification purposes. First, it is easy to see that when all $\psi_{kl,jh}$-parameters are 0, the conditional probabilities simplify to standard multinomial logit probabilities where the K choices are independent. Hence, to identify the parameters we have to impose the standard identification restrictions of the multinomial logit model, that is, $\alpha_{k,1} = 0$ and $\beta_{k,1} = 0$ for all k. Second, the conditional probabilities should be compatible with one another to be able to form a proper joint distribution. That is, the association of choice option j for choice k with choice option h for choice l should be the same as the opposite relation. Formally, we have to impose $\psi_{kl,jh} = \psi_{lk,hj}$ for symmetry. Finally, as utility differences determine choice, we cannot identify all association parameters. That is, equal changes in all associations would not yield any difference. Therefore, we impose that $\psi_{kl,j1} = \psi_{kl,1h} = 0$ for all j and h without loss of generality. Note that it is possible to impose other identification restrictions. In fact, in a large unrestricted MNL which describes the $\prod_{k=1}^{K} J_k$ possible realizations of the random variable Y_i, we can identify $(\prod_{k=1}^{K} J_k) - 1$ intercept parameters and the same number of β parameters. This implies that -together with standard univariate multinomial logit restrictions- our restrictions on the β parameters are sufficient for identification as long as one set of choices contains a full set of restricted β parameters. For α and ψ parameters the restrictions are necessary for $K = 2$ and sufficient for $K > 2$. Our choices to identify the parameters are however motivated by: (i) they are a straightforward extension of parameter identification in the multivariate binomial choice model of Russell and Petersen (2000); (ii) they are universal, that is, can be applied for all possible values of K and J_k and; (iii) they yield direct interpretations of the association parameters via odds ratios. Section 3.2.1 describes the need and implications of these identification restrictions.

The discussion can easily be changed to a multivariate conditional logit specification where the explanatory variables instead of parameters vary over alternative choices. Hence, the exponent in (3.2) then writes

$$Z_{ik,j} \quad = \quad \alpha_{k,j} + W_{ik,j}\gamma_k + \sum_{l \neq k} \psi_{kl,jy_{il}}, \tag{3.8}$$

where $W_{ik,j}$ denotes the value of the explanatory variables which now differs over i, k and j and γ_k denotes the corresponding parameter vector. The joint probabilities are then given by (3.4) with

$$\mu_{y_i} \quad = \quad \sum_{k=1}^{K} \left(\alpha_{k,y_{ik}} + W_{ik,y_{ik}}\gamma_k + \sum_{l>k} \psi_{kl,y_{ik}y_{il}} \right). \tag{3.9}$$

The proof directly follows from the proof for the MV-MNL specification in Appendix 3.A.1.[1]

Finally, the model can be extended with individual-specific association parameters by replacing the expression for $\psi_{kl,jh}$ in (3.5) by

$$\psi_{i,kl,jh} = \xi_{kl,jh} + X_i\delta_{kl,jh}, \tag{3.10}$$

where $\xi_{kl,jh}$ and $\delta_{kl,jh}$ are additional parameters. The association between decisions j and h now depends on individual characteristics X_i.

3.2.1 A Bivariate Trinomial Logit Model

To illustrate the properties of the proposed multivariate multinomial logit model and the need for identification restrictions we consider a bivariate trinomial logit specification. Hence, we have $K = 2$ and $J_1 = J_2 = 3$. The conditional probabilities with the proper identification restrictions imposed are defined as

$$\begin{aligned}
\Pr[Y_{ik} = 1|y_{il}, X_i] &\propto 1 \\
\Pr[Y_{ik} = 2|y_{il}, X_i] &\propto \exp(\alpha_{k,2} + X_i\beta_{k,2} + \psi_{kl,2y_{il}}) \\
\Pr[Y_{ik} = 3|y_{il}, X_i] &\propto \exp(\alpha_{k,3} + X_i\beta_{k,3} + \psi_{kl,3y_{il}}).
\end{aligned} \tag{3.11}$$

These conditional probabilities imply the following 9 choice probabilities:

$$\begin{aligned}
\Pr[Y_i = (1,1)'|X_i] &\propto 1 \\
\Pr[Y_i = (1,2)'|X_i] &\propto \exp(\alpha_{2,2} + X_i\beta_{2,2}) \\
\Pr[Y_i = (1,3)'|X_i] &\propto \exp(\alpha_{2,3} + X_i\beta_{2,3}) \\
\Pr[Y_i = (2,1)'|X_i] &\propto \exp(\alpha_{1,2} + X_i\beta_{1,2}) \\
\Pr[Y_i = (2,2)'|X_i] &\propto \exp(\alpha_{1,2} + \alpha_{2,2} + X_i(\beta_{1,2} + \beta_{2,2}) + \psi_{12,22}) \\
\Pr[Y_i = (2,3)'|X_i] &\propto \exp(\alpha_{1,2} + \alpha_{2,3} + X_i(\beta_{1,2} + \beta_{2,3}) + \psi_{12,23}) \\
\Pr[Y_i = (3,1)'|X_i] &\propto \exp(\alpha_{1,3} + X_i\beta_{1,3}) \\
\Pr[Y_i = (3,2)'|X_i] &\propto \exp(\alpha_{1,3} + \alpha_{2,2} + X_i(\beta_{1,3} + \beta_{2,2}) + \psi_{12,32}) \\
\Pr[Y_i = (3,3)'|X_i] &\propto \exp(\alpha_{1,3} + \alpha_{2,3} + X_i(\beta_{1,3} + \beta_{2,3}) + \psi_{12,33}).
\end{aligned} \tag{3.12}$$

[1] The proof requires that $Z_{ik,1} = 0$ which does not hold for this specification. We can however rewrite the model such that $Z_{ik,j} = \alpha_{k,j} + (W_{ik,j} - W_{ik,1})\gamma_k + \sum_{l \neq k} \psi_{kl,jy_{il}}$ with $Z_{ik,1} = 0$ such that the proof is similar as in Appendix 3.A.1.

As we have 9 probabilities we can only identify 8 different intercept parameters. The imposed identification restrictions result in exactly 4 α-parameters and 4 ψ-parameters and thus cause identifiability. It is easy to see that imposing $\psi_{12,22} = \psi_{12,23} = \psi_{12,32} = \psi_{12,33} = 0$ implies that the joint probabilities can be written as the product of two independent multinomial logit probabilities. Furthermore, we see that

$$\psi_{12,jh} = \ln\left(\frac{\Pr[Y_i = (j,h)'|X_i]\,\Pr[Y_i = (1,1)'|X_i]}{\Pr[Y_i = (j,1)'|X_i]\,\Pr[Y_i = (1,h)'|X_i]}\right). \tag{3.13}$$

Hence, a positive value of $\psi_{12,jh}$ implies positive association between choosing j for choice 1 and h for choice 2.

3.2.2 Parameter Inference

To estimate the parameters of the multivariate binary logit model Russell and Petersen (2000) suggest to use maximum likelihood using a log-likelihood function based on the joint probabilities, that is

$$\ell(\theta; y) = \sum_{i=1}^{N} \ln \Pr[Y_i = y_i|X_i], \tag{3.14}$$

where $\Pr[Y_i = y_i|X_i]$ is given in (3.4) and where θ summarizes the model parameters.

The same approach is of course possible for our MV-MNL specification. The disadvantage is however that the computation of these joint probabilities may be a burden if the dimensions of the logit specification are large. For example, for $K = 10$ and $J_k = 5$ for all k we have to take the sum of 5^{10} different terms in the denominator of the joint probabilities. The outcome space of the multivariate multinomial random variable rapidly grows large and the computation time thereby increases exponentially with the number of choices. Further, numerical problems emerge as probabilities get small (in the given example $1/5^{10}$ on average).

To avoid these numerical problems, we propose another estimation approach based on the ideas in Chapter 2 and Bel et al. (2014) for the MVL specification. Bel et al. (2014) propose to use a composite likelihood approach (Lindsay, 1988) using all conditional probabilities in the likelihood specification (Molenberghs and Verbeke, 2005, chapter 12) instead of the joint probabilities in (3.4). The resulting composite conditional likelihood [CCL] representation only uses conditional probabilities and hence it avoids summation over the complete outcome space. It thereby also avoids the computation of numerically small probabilities. It can be shown that the CCL approach provides consistent estimators (Varin et al., 2011) but at the cost of loss in efficiency.

The conditional probabilities in (3.2) lead to the composite log-likelihood function of the MV-MNL specification, that is

$$\begin{aligned} \ell^c(\theta; y) &= \sum_{i=1}^{N} \ell^c(\theta; y_i) \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K} \ell^c(\theta; y_{ik}) \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K} \ln P[Y_{ik} = y_{ik} | y_{il} \text{ for } l \neq k, X_i]. \end{aligned} \tag{3.15}$$

The estimator $\hat{\theta}$ which follows from maximizing (3.16) is consistent. Varin et al. (2011) show that standard errors in CCL can be computed using the Godambe (1960) information matrix, which has a sandwich form and writes

$$G_{\hat{\theta}} = H_{\hat{\theta}} J_{\hat{\theta}}^{-1} H_{\hat{\theta}} \tag{3.16}$$

with

$$H_{\hat{\theta}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} \nabla\ell^c(\hat{\theta}; y_{ik}) \nabla\ell^{c'}(\hat{\theta}; y_{ik}) \tag{3.17}$$

and

$$J_{\hat{\theta}} = \frac{1}{N}\sum_{i=1}^{N} \nabla\ell^c(\hat{\theta}; y_i) \nabla\ell^{c'}(\hat{\theta}; y_i). \tag{3.18}$$

where $\nabla\ell^c(\hat{\theta}; y_{ik})$ and $\nabla\ell^c(\hat{\theta}; y_i)$ denote the first-order derivatives of the corresponding log-likelihood contributions in (3.16). The covariance matrix of the parameter estimates is then given by $(-G_{\hat{\theta}})^{-1}$.

To test for independence in the multinomial decisions one can use a Likelihood Ratio [*LR*] statistic for the restriction that the association parameters ψ equal 0. This *LR*-statistic does not have a standard distribution when the CCL estimation approach is used. Based on results by Satterthwaite (1946) and Kent (1982), Varin et al. (2011) propose to use an adjusted *LR*-statistic which for our test of independence boils down to

$$LR = \frac{\nu}{Q\bar{\lambda}} 2\left(\ell^c(\hat{\theta}; y) - \ell^c(\hat{\alpha}, \hat{\beta}; y)\right), \tag{3.19}$$

where $\ell^c(\hat{\theta}; y)$ is the value of the CCL evaluated in the estimate under the alternative hypothesis and $\ell^c(\hat{\alpha}, \hat{\beta}; y)$ the value of the CCL evaluated in the estimate under the null and where Q is the number of ψ parameters. This *LR*-statistics is asymptotically $\chi^2(\nu)$ distributed with

$$\nu = \frac{\left(\sum_{q=1}^{Q} \lambda_q\right)^2}{\sum_{q=1}^{Q} \lambda_q^2}, \tag{3.20}$$

where $\lambda_1, \ldots, \lambda_Q$ are eigenvalues of $(G_\psi (H^{-1})_\psi)^{-1}$ with G_ψ the $Q \times Q$ sub-matrix of the Godambe information matrix corresponding to ψ. Moreover, $\bar{\lambda}$ denotes the average of the eigenvalues.

Although the composite conditional likelihood does not correspond to the true likelihood function, it still takes the correlation between choice decisions in the multivariate multinomial logit model into account. The advantage over the full multinomial representation in (3.4) is that CCL avoids the large summation in the denominator. It is therefore possible to compute CCL even if there is a large number of choices and alternatives. Nonetheless, since the composite instead of the true likelihood function is used, the estimator is not efficient. Chapter 2 and Bel et al. (2014) show that the loss in efficiency is quite small for MVL models. In the next subsection we conduct a small Monte Carlo study to analyze the efficiency loss for the MV-MNL specification.

3.2.3 Monte Carlo Study

In this section we conduct a Monte Carlo study to investigate the properties of the composite likelihood estimator for the parameters of a multivariate multinomial logit specification. We focus on potential small sample bias and loss in efficiency caused by using the composite instead of the exact log-likelihood specification in the estimation procedure. Finally, we check whether the normal distribution can be used to approximate the small sample distribution of the CCL estimator.

For our Monte Carlo study we consider the MV-MNL specification (3.4) with (3.5). The number of choices K is fixed to 3 and the number of choice alternatives per choice are $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$. We consider a relatively small sample size $N = 250$ and a large sample $N = 5000$. As explanatory variables X_i we take two positively correlated random variables; one continuous and one discrete. Both variables are drawn from a bivariate normal distribution with variances 0.25 and correlation 0.75. The second variable is made discrete based on a zero threshold. The parameters of our data generating process [DGP] are chosen such that there is an unequal distribution over the choice alternatives but still substantial choice probabilities for every choice combination, see Tables 3.1 and 3.2 for the values of our DGP-parameters.

Table 3.1: Mean and RMSE of the estimator for the MV-MNL model parameters based on a Monte Carlo study with $N = 250$ (10000 replications)[a]

		$k=1$				$k=2$				$k=3$		
		θ	$\hat{\theta}$	RMSE		θ	$\hat{\theta}$	RMSE		θ	$\hat{\theta}$	RMSE
	$\alpha_{1,2}$	0.15	0.165	0.684	$\alpha_{2,2}$	0.15	0.172	0.758	$\alpha_{3,2}$	0.150	0.164	0.827
	$\alpha_{1,3}$	0.25	0.278	0.690	$\alpha_{2,3}$	0.25	0.290	0.753	$\alpha_{3,3}$	0.250	0.262	0.830
					$\alpha_{2,4}$	0.375	0.426	0.757	$\alpha_{3,4}$	0.375	0.424	0.781
									$\alpha_{3,5}$	0.475	0.540	0.777
X_1	$\beta_{1,2}$	1.05	1.110	0.525	$\beta_{2,2}$	1.05	1.122	0.661	$\beta_{3,2}$	1.05	1.138	0.723
	$\beta_{1,3}$	1.45	1.533	0.526	$\beta_{2,3}$	1.45	1.545	0.650	$\beta_{3,3}$	1.45	1.556	0.710
					$\beta_{2,4}$	1.75	1.864	0.650	$\beta_{3,4}$	1.75	1.878	0.677
									$\beta_{3,5}$	1.95	2.083	0.674
X_2	$\beta_{1,2}$	0.25	0.258	0.464	$\beta_{2,2}$	0.25	0.272	0.635	$\beta_{3,2}$	0.25	0.274	0.702
	$\beta_{1,3}$	0.45	0.479	0.452	$\beta_{2,3}$	0.45	0.490	0.602	$\beta_{3,3}$	0.45	0.486	0.671
					$\beta_{2,4}$	0.65	0.696	0.594	$\beta_{3,4}$	0.65	0.697	0.637
									$\beta_{3,5}$	0.80	0.850	0.628

[a] The DGP is given in Section 3.2.3 with $K = 3$ and $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$.

Table 3.2: Mean and RMSE of the estimator for the MV-MNL model parameters based on a Monte Carlo study with $N = 5000$ (10000 replications)[a]

		$k=1$				$k=2$				$k=3$		
		θ	$\hat{\theta}$	RMSE		θ	$\hat{\theta}$	RMSE		θ	$\hat{\theta}$	RMSE
	$\alpha_{1,2}$	0.15	0.141	0.131	$\alpha_{2,2}$	0.150	0.143	0.154	$\alpha_{3,2}$	0.150	0.138	0.166
	$\alpha_{1,3}$	0.25	0.243	0.137	$\alpha_{2,3}$	0.250	0.242	0.149	$\alpha_{3,3}$	0.250	0.241	0.167
					$\alpha_{2,4}$	0.375	0.375	0.145	$\alpha_{3,4}$	0.375	0.367	0.149
									$\alpha_{3,5}$	0.475	0.467	0.158
X_1	$\beta_{1,2}$	1.05	1.049	0.107	$\beta_{2,2}$	1.05	1.062	0.129	$\beta_{3,2}$	1.05	1.046	0.144
	$\beta_{1,3}$	1.45	1.447	0.101	$\beta_{2,3}$	1.45	1.462	0.128	$\beta_{3,3}$	1.45	1.454	0.141
					$\beta_{2,4}$	1.75	1.762	0.130	$\beta_{3,4}$	1.75	1.752	0.138
									$\beta_{3,5}$	1.95	1.953	0.136
X_2	$\beta_{1,2}$	0.25	0.254	0.098	$\beta_{2,2}$	0.25	0.243	0.122	$\beta_{3,2}$	0.25	0.258	0.147
	$\beta_{1,3}$	0.45	0.455	0.099	$\beta_{2,3}$	0.45	0.443	0.120	$\beta_{3,3}$	0.45	0.455	0.137
					$\beta_{2,4}$	0.65	0.644	0.121	$\beta_{3,4}$	0.65	0.658	0.132
									$\beta_{3,5}$	0.80	0.806	0.133

[a] The DGP is given in Section 3.2.3 with $K = 3$ and $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$.

Table 3.3: Mean and RMSE of the estimator for the association parameters based on a Monte Carlo study with $N = 250$ (10000 replications)[a]

	$k = 2$											
		2			3			4				
$k = 1$	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE			
2	0.475	0.513	0.612	0.250	0.267	0.598	0	-0.001	0.601			
3	0.250	0.271	0.617	0.475	0.492	0.597	0.25	0.249	0.588			
	$k = 3$											
		2			3			4			5	
$k = 1$	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE
2	-0.375	-0.409	0.704	-0.15	-0.155	0.663	0	-0.002	0.636	0.15	0.152	0.646
3	-0.150	-0.168	0.672	-0.375	-0.389	0.668	-0.15	-0.166	0.636	0	-0.010	0.642
	$k = 3$											
		2			3			4			5	
$k = 2$	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE
2	0.475	0.512	0.814	0.250	0.289	0.827	0	0.008	0.799	-0.25	-0.262	0.780
3	0.250	0.278	0.833	0.475	0.529	0.808	0.25	0.268	0.772	0	0.004	0.756
4	0	0.011	0.862	0.250	0.288	0.834	0.475	0.512	0.781	0.25	0.265	0.761

[a] The DGP is given in Section 3.2.3 with $K = 3$ and $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$.

Table 3.4: Mean and RMSE of the estimator for the association parameters based on a Monte Carlo study with $N = 5000$ (10000 replications)[a]

	$k = 2$											
		2			3			4				
$k = 1$	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE			
2	0.475	0.483	0.122	0.250	0.258	0.120	0	0.001	0.122			
3	0.250	0.256	0.131	0.475	0.485	0.122	0.25	0.252	0.119			
	$k = 3$											
		2			3			4			5	
$k = 1$	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE
2	-0.375	-0.374	0.147	-0.150	-0.151	0.136	0	0.003	0.129	0.15	0.153	0.134
3	-0.150	-0.143	0.137	-0.375	-0.371	0.138	-0.15	-0.146	0.126	0	0.001	0.130
	$k = 3$											
		2			3			4			5	
$k = 2$	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE
2	0.475	0.482	0.164	0.25	0.259	0.163	0	0.004	0.160	-0.25	-0.237	0.155
3	0.250	0.256	0.166	0.475	0.482	0.152	0.25	0.256	0.153	0	0.008	0.150
4	0	0.001	0.161	0.25	0.255	0.162	0.475	0.477	0.157	0.25	0.257	0.149

[a] The DGP is given in Section 3.2.3 with $K = 3$ and $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$.

Tables 3.1 to 3.4 display the mean and root mean squared error [RMSE] of the CCL estimator. The final two tables show that for $N = 5000$ the bias in the estimator is quite small. For a smaller sample size $N = 250$, the deviation from the DGP parameters is larger. Unreported results show that the bias is almost the same as the bias in a regular maximum likelihood approach.[2] The RMSE shows that there is a large variance of the estimator for small sample sizes. This is not a surprise as we in fact try to estimate the parameters of an MNL model with $3 \times 4 \times 5 = 60$ choice alternatives using only 250 observations.

To analyze the loss in efficiency between CCL and the regular likelihood approach, we consider the ratio of the RMSEs of both approaches. Table 3.5 shows a selection of these ratios which is based on best and worst cases. The ratios are close to 1 and hence the loss in efficiency is rather limited even in small samples. For example, for the largest difference, CCL is only 1.3 percent worse in RMSE than regular ML. Hence, CCL seems to be a valid alternative for maximum likelihood to estimate the parameters of an MV-MNL model. The small-sample bias is similar and the loss in efficiency is very small.

Table 3.5: Relative RMSE of the maximum CCL and the regular ML estimator[a]

	Sample size	
Parameter	250	5000
$\alpha_{1,2}$	1.007	1.003
$\alpha_{2,3}$	1.007	1.003
$\alpha_{3,4}$	1.007	1.000
$\beta_{1,3}$	1.008	1.000
	1.002	1.000
$\beta_{2,4}$	1.013	1.000
	1.004	1.001
$\beta_{3,5}$	1.013	1.003
	1.002	1.001
$\psi_{12,22}$	1.005	1.001
$\psi_{13,33}$	1.005	1.000
$\psi_{23,44}$	1.006	1.002

[a] We only report results for a subset of parameters. The results for the other parameters are similar and available upon request.

[2] Detailed results are available upon request.

Table 3.6: Empirical size of the distribution of the estimators based on a Monte Carlo study with $N = 250$ (10000 replications)[a]

Theoretical	0.025	0.05	0.95	0.975
$\alpha_{1,2}$	0.028	0.055	0.951	0.976
$\alpha_{2,3}$	0.027	0.053	0.954	0.977
$\alpha_{3,4}$	0.022	0.048	0.957	0.980
$\beta_{1,3}$	0.031	0.058	0.954	0.979
	0.027	0.053	0.953	0.978
$\beta_{2,4}$	0.031	0.056	0.910	0.957
	0.029	0.055	0.908	0.958
$\beta_{3,5}$	0.029	0.054	0.962	0.982
	0.031	0.055	0.961	0.982
$\psi_{12,22}$	0.034	0.059	0.952	0.977
$\psi_{13,33}$	0.026	0.054	0.943	0.971
$\psi_{23,44}$	0.029	0.056	0.954	0.979

[a] We only report results for a subset of parameters. The results for the other parameters are similar and available upon request.

Apart from bias and efficiency, we also consider the validity of using a normal distribution for testing for significance of the parameters. Table 3.6 displays the empirical size of the t-tests for $N = 250$ for both tails of t-statistics. The table shows that even for $N = 250$ size distortions are rather small. For example, a theoretical 90 percent confidence interval for $\psi_{13,33}$ turns out to have coverage of 88.8 percent. This size distortion is acceptable.

In sum, the simulation study shows that the composite likelihood estimator has similar small-sample biases as the maximum likelihood estimator and that efficiency losses are limited. Inference based on t-statistics seems to be valid even in relatively small samples. Because of the advantages of CCL over ML when dimensions increase, CCL is a good alternative for the estimation of parameters in a multivariate multinomial logit specification. In Section 3.4.1, we will use the CCL approach in a small application.

3.3 A Panel Specification

The MV-MNL model can easily be extended to a fixed-effects panel data specification. Let Y_{it} denote the K-dimensional random variable describing the joint set of choices for individual $i = 1, \ldots, N$ at time $t = 1, \ldots, T$ and let $Y_{itk} = j$ if individual i chooses $j = 1, \ldots, J_k$ for the k-th choice at time t. The choice probabilities are given by

$$\Pr[Y_{it} = y_{it}|X_{it}] \quad = \quad \frac{\exp(\mu_{y_{it}})}{\sum_{s_{it} \in S} \exp(\mu_{s_{it}})}, \tag{3.21}$$

where y_{it} is a possible realization from the outcome space S and where $\mu_{y_{it}}$ is defined as

$$\mu_{y_{it}} = \sum_{k=1}^{K} \left(\alpha_{ik,y_{itk}} + X_{it}\beta_{k,y_{itk}} + \sum_{l>k} \psi_{ikl,y_{itk}y_{itl}} \right). \tag{3.22}$$

Hence, both the intercepts and the association parameters are individual specific. A special case of the model is where the association parameters are pooled across the individuals in which case we replace $\psi_{ikl,y_{itk}y_{itl}}$ in (3.22) by $\psi_{kl,y_{itk}y_{itl}}$.

3.3.1 Parameter Inference

In practice the number of cross sections is usually limited and hence parameter estimation suffers from the incidental parameter problem. To solve this, we follow Chamberlain (1980) and Lee (2002, Chapter 6) who condition on a sufficient statistic which eliminates the fixed effects from the model specification. We extend the solution of Chamberlain (1980) for a univariate panel MNL model to our multivariate multinomial setting in (3.21). The appropriate sufficient statistics are given by

$$v_{i,s}^{(1)} = \sum_{t=1}^{T} I[Y_{it} = s] = c_{i,s} \qquad \forall s \in S, \tag{3.23}$$

where $c_{i,s}$ represents the number of times the combination of choices s occurs for individual i. Thus, only the alternatives containing the same choice sets over time as observed for individual i are used in the logit specification. That is, only the permutations of choices of individual i over time are taken into account. Since no permutations can be made for individuals where no change takes place over time, these observations are not of interest and discarded. Appendix 3.A.2 shows that the choice probabilities conditionally on these sufficient statistics are given by

$$\Pr[Y_i = y_i | v_i^{(1)}, X_i] = \frac{\exp\left(\sum_{t=1}^{T}\sum_{k=1}^{K} X_{it}\beta_{k,y_{itk}}\right)}{\sum_{d_i \in B_i} \exp\left(\sum_{t=1}^{T_i}\sum_{k=1}^{K} X_{it}\beta_{k,d_{itk}}\right)}, \tag{3.24}$$

where B_i is the set of alternatives for which $v_i^{(1)}$ holds. Hence, the individual-specific parameters (intercepts and association parameters) are removed from the probabilities and the β-parameters can be estimated consistently using a log-likelihood function where we condition on the sufficient statistics. Note that this approach only works if X_{it} does not depend on lagged dependent variables.

In case the association parameters are of core interest, these should not be discarded from the specification. Therefore, we make $\psi_{kl,jh}$ not individual-specific and we have to consider other sufficient statistics

$$v_{i,k,j}^{(2)} = \sum_{t=1}^{T} I[Y_{itk} = j] = c_{i,k,j} \qquad \forall k, j, \tag{3.25}$$

where $c_{i,k,j}$ now represents the number of times that individual i chooses option j for the k-th choice. Appendix 3.A.2 shows that when we condition on these sufficient statistics the choice probabilities are given by

$$\Pr[Y_i = y_i | v_i^{(2)}, X_i] = \frac{\exp\left(\sum_{t=1}^{T}\sum_{k=1}^{K}\left(X_{it}\beta_{k,y_{itk}} + \sum_{l>k}\psi_{kl,y_{itk}y_{itl}}\right)\right)}{\sum_{d_i \in B_i}\exp\left(\sum_{t=1}^{T}\sum_{k=1}^{K}\left(X_{it}\beta_{k,d_{itk}} + \sum_{l>k}\psi_{kl,d_{itk}d_{itl}}\right)\right)}, \tag{3.26}$$

where $\psi_{kl,jh}$ does not drop out since the combination of choices may differ over the alternatives in set B_i where $v_i^{(2)}$ holds. Hence, we now can find estimates of both $\beta_{k,j}$ and the association parameters $\psi_{kl,jh}$ describing the relation of the choices in the multivariate multinomial logit specification. Again this approach is only valid if X_{it} does not contain lagged dependent variables.

The disadvantage of using the log-likelihood function conditional on the sufficient statistics for parameter estimation is again the sum over many alternatives in the denominator of the choice probabilities which causes numerical problems. In Appendix 3.A.3 we however show that the Composite Likelihood method can also be applied in a panel data setting. We thereby avoid the extensive sum and make parameter estimation of MV-MNL models feasible in a panel context.

In the next section we illustrate the possibilities of the MV-MNL model by applications of its panel version discussed in this section and its cross-sectional counterpart from Section 3.2 to household panel scanner data and a survey on life satisfaction, respectively.

3.4 Application

This section considers two illustrations of our newly proposed MV-MNL model. First, we apply the model on cross sectional survey data on satisfaction. Satisfaction is measured at an ordinal scale and satisfaction levels on different items are likely to be correlated. Hence, the MV-MNL model specification from Section 3.2 and the CCL estimation procedure from Section 3.2.2 can be used. Second, we investigate the product choice of canned tuna fish in a household panel scanner data set. Various multinomial choices on the characteristics of canned tuna fish are made. As these decisions are made simultaneously (and not in some natural ordering assumed in nested logit models) the model presented in Section 3.3 is highly applicable.

3.4.1 Survey Data on Satisfaction

To illustrate the MV-MNL model discussed in Section 3.2, we consider modeling satisfaction of 2012 Dutch respondents to an extensive survey from 2004.[3] Satisfaction is represented by 5 ordinal dependent variables: Satisfaction about *Life*, *Income*, the *Social* security system, *Democracy* and the *Government*. For *Life*, *Income* and *Democracy* respondents can be *Satisfied*, *Unsatisfied* or *In between*. *Social* and *Government* have two options: either the respondent is *Satisfied* or (s)he is *Unsatisfied*. The base category is *Satisfied* such that a positive β-parameter indicates less satisfaction if x_i is large and positive. To describe relations in satisfaction level we consider the MV-MNL model of Section 3.2 with $K = 5$, $J_1 = J_2 = J_4 = 3$ and $J_3 = J_5 = 2$. As explanatory variables we have *Gender*, *Age*, *Unemployment*, (self-reported) *Health* status, *Religion*, *Political interest* and *Income*.

Since our dependent variables are ordered multinomial variables we opt for a stereotype logit specification (Anderson, 1984). That is, we adjust our model specification in (3.5) such that the parameter estimates are restricted to be monotonically increasing or decreasing over the choice options. Formally, we change (3.5) into

$$\mu_{y_i} \quad = \quad \sum_{k=1}^{K} \left(\alpha_{k,y_{ik}} + \phi_{k,y_{ik}}(X_i \beta_k) + \sum_{l>k} \psi_{kl,y_{ik}y_{il}} \right), \tag{3.27}$$

where $0 = \phi_{k,1} < \cdots < \phi_{k,J_k} = 1$ for ordering and identification purposes. This addition to the model specification does not change the general set-up of our proposed estimation procedures.

We use the composite likelihood method to estimate the model parameters in (3.27). First, we test for independence among the five satisfaction levels. The *LR*-statistic for the restriction that all $\psi_{kl,jh}$ are 0 equals 1808.94. Since the degrees of freedom of the approximate χ^2-distribution is 50.44, independence is clearly rejected. Hence, we find positive support for association between the levels of satisfaction under consideration.

Tables 3.7 and 3.8 display the parameter estimates and estimated standard errors from the CCL method. The majority of respondents is satisfied about life, income, social security and the government, which results in negative estimates of the choice-specific intercepts. The effect for *Government* is modest. The positive estimate of the α_2 intercept shows less baseline satisfaction on *Democracy*.

Several relations between the explanatory variables and satisfaction are found. Note that since *Satisfied* is the base category, a negative β-parameter indicates that the probability to be satisfied gets larger when x_i increases. For example, individuals with low (high) self-reported *Health* status are

[3]This data is freely available at the website of the The Netherlands Institute for Social Research: http://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Culturele_veranderingen_in_Nederland_CV

Table 3.7: Parameter estimates of the MV-MNL model for a cross-sectional survey on satisfaction[ab] (standard errors in parentheses)

	Life		Income		Social Sec.		Democracy		Government	
α_2[c]	-0.526	(0.238)	-1.096	(0.226)	-2.070	(0.270)	1.246	(0.228)	-0.667	(0.233)
α_3	-2.125	(0.351)	-2.307	(0.351)	–		-0.811	(0.341)	–	
ϕ_2	0.737	(0.073)	0.519	(0.062)	–		0.625	(0.058)	–	
gender	-0.266	(0.128)	-0.520	(0.143)	-0.023	(0.107)	0.870	(0.187)	0.189	(0.103)
younger than 35	-0.231	(0.168)	0.342	(0.175)	-0.243	(0.133)	0.010	(0.241)	-0.291	(0.130)
older than 54	-0.348	(0.175)	-1.023	(0.189)	-0.578	(0.146)	0.065	(0.257)	-0.127	(0.143)
unemployment	0.257	(0.165)	0.049	(0.171)	0.368	(0.134)	-0.304	(0.247)	-0.226	(0.135)
low health status	0.888	(0.159)	0.531	(0.171)	0.286	(0.132)	-0.052	(0.233)	-0.129	(0.129)
high health status	-1.521	(0.196)	0.051	(0.200)	-0.063	(0.144)	-0.916	(0.232)	-0.175	(0.133)
religion	-0.041	(0.138)	-0.209	(0.148)	-0.174	(0.112)	0.016	(0.194)	-0.335	(0.110)
low political interest	0.159	(0.137)	-0.105	(0.150)	0.023	(0.115)	0.982	(0.206)	-0.117	(0.111)
high political interest	-0.015	(0.191)	-0.557	(0.218)	0.083	(0.160)	-0.289	(0.251)	0.526	(0.161)
low income	0.104	(0.151)	0.970	(0.162)	-0.029	(0.127)	0.273	(0.223)	-0.158	(0.125)
high income	-0.677	(0.166)	-0.924	(0.202)	-0.002	(0.139)	-0.204	(0.235)	0.044	(0.129)

[a] The data is from an extensive survey on cultural development from 2004 by the The Netherlands Institute for Social Research.
[b] The model specification is described in (3.27).
[c] Higher subscripts indicate lower satisfaction (the baseline is ’satisfied’).

ceteris paribus more likely to report low (high) satisfaction about life. Furthermore, both women and respondents of higher age are more satisfied about their income than respectively men and respondents of average age. Unemployed respondents are more likely to report low satisfaction on the social security system. Respondents with low political interest tend to have ceteris paribus less satisfaction on democracy. Finally, religious respondents report to be more satisfied about the (at that time Christian-Liberal) government than nonreligious respondents.

The estimates of the association parameters ψ in Table 3.8 indicate the relation between reported satisfaction levels for the five dependent variables. Clear interpretations can be given. All parameter values that are significantly different from 0 are positive. That is, there is a positive relation between the reported satisfaction levels of respondents. For example, $\phi_{Life\ Income,33}$ indicates that respondents who report *Unsatisfied* on *Income* are likely also to be unsatisfied about life. Respondents unsatisfied about the social security system are more likely also to be unsatisfied about both *Democracy* and *Government*. This can be explained by the Labor party ending second in the previous elections with 27% of the votes but not being in charge.

Table 3.8: Estimates of the association parameters of the MV-MNL model for a cross-section survey on satisfaction[a] (standard errors in parentheses)

	Life				Income				Social Sec.		Democracy			
	2		3		2		3		2		2		3	
Income														
2	0.327	(0.116)	0.454	(0.184)										
3	0.740	(0.184)	1.948	(0.217)										
Social														
2	0.179	(0.120)	0.342	(0.166)	0.430	(0.120)	0.887	(0.155)						
Democracy														
2	0.464	(0.196)	0.280	(0.292)	0.348	(0.203)	-0.119	(0.270)	0.361	(0.214)				
3	0.556	(0.217)	0.408	(0.313)	0.444	(0.224)	0.340	(0.288)	1.058	(0.225)				
Government														
2	0.125	(0.114)	0.599	(0.171)	0.377	(0.114)	0.642	(0.163)	0.951	(0.113)	0.261	(0.178)	1.445	(0.200)

[a] As the association parameters are symmetric only the lower triangular matrix is given.

3.4.2 Household Panel Scanner Data

To illustrate the MV-MNL model in a panel data setting we consider product choices of canned tuna in 21 supermarkets belonging to 4 chains for 1092 individuals during the period 1986(week 25)–1987(week 23) in Springfield, Missouri.[4] For each household we take the first 5 purchases in the sample and hence $T = 5$. The product choice of canned tuna concerns choosing from four characteristics: *Brand* (Chicken of the Sea, Star-Kist, CTL), whether it is *Oil*-based or not, whether it is a *Light*-product or not and *Volume* of the can. There are three choice options for *Brand* and two for the remaining characteristics. We assume that individuals make choices for these characteristics simultaneously and hence the multivariate multinomial choice model of Section 3.3 is applicable. That is, we consider a panel data MV-MNL model with $K = 4$, $J_1 = 3$ and $J_2 = \cdots = J_4 = 2$ with $N = 1092$ and $T = 5$. The base category for each of the 4 choices is taken to be the characteristic of the market leader.

As explanatory variables for product choice, we take the product-specific marketing-mix variables *Price* of the product, *Display* and *Feature*. Hence, (3.22) becomes

$$\mu_{y_{it}} = \sum_{k=1}^{K} \left(\alpha_{ik,y_{itk}} + W_{ity_{it}}\gamma + \sum_{l>k} \psi_{ikl,y_{itk}y_{itl}} \right), \tag{3.28}$$

where $W_{ity_{it}}$ are now choice-specific variables. We consider two model specifications. In the first specification the ψ-parameters are individual-specific. The second specification contains ψ-parameters for all households. Hence, we respectively use sufficient statistics $v_{i,s}^{(1)}$ and $v_{i,s}^{(2)}$.

Table 3.9 displays the parameter estimates and estimated standard errors from the model specification with individual-specific association parameters. Parameter estimates are obtained using a likelihood approach using (3.23) as sufficient statistic. Hence, the individual-specific association parameters ψ are not estimated.

[4] This data set is from the ERIM Database and publicly available at http://research.chicagobooth.edu/kilts/marketing-databases/erim/erim-dataset

Table 3.9: Parameter estimates of the MV-MNL model for a household panel scanner data set on canned tuna product choice (standard errors in parentheses)[a]

	$\hat{\gamma}$	s.e.
Price	-0.366	(0.017)
Display	0.888	(0.117)
Feature	1.416	(0.087)

[a] Results are obtained using sufficient statistics (3.23).

To interpret the parameter estimates, we opt for conditional marginal effects

$$\frac{\partial \Pr[Y_{itk} = j | y_{itl} \text{ for } l \neq k, X_{it}, W_{ity_{it}}]}{\partial w_{ity_{it}}} = \gamma \Pr[Y_{itk} = j | y_{itl} \text{ for } l \neq k, X_{it}, W_{ity_{it}}] \times \\ (1 - \Pr[Y_{itk} = j | y_{itl} \text{ for } l \neq k, X_{it}, W_{ity_{it}}]) \, . \quad (3.29)$$

By averaging these over y_{itl} ($l \neq k$) and the explanatory variables, that is,

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \Pr[Y_{itk} = j | y_{itl} \text{ for } l \neq k, W]}{\partial w} \quad (3.30)$$

we obtain an estimate for the average marginal effects. Table 3.10 reports these effects. An increase in *Price* leads to a decrease in the probability for each product characteristic. Equation (3.29) shows that the maximum marginal effect takes place when $\Pr[Y_{itk} = j | X_{it}, W_{ity_{it}}] = 0.5$ and equals $1/4$ of the parameter estimate in Table 3.9. The effect is on average larger for the probability to buy large *Volume* products and relative small for water-based canned tuna. Both increases in *Display* and *Feature* have a positive effect on the probability for each product characteristic, where the effect of *Feature* is larger. A product with characteristics *Brand* Star-Kist, *Oil*-based, *Light* and large volume would especially gain from advertisements, given the relatively large marginal effects.

Table 3.10: Average marginal effects of 1 unit change in explanatory variables to marginal probabilities for the tuna product characteristic choices (standard deviations in parentheses)

	Brand						Oil				Light				Volume			
	Ckn of Sea		Star-Kist		CTL		No		Yes		No		Yes		Small		Large	
Price	-0.025	(0.025)	-0.027	(0.022)	-0.022	(0.027)	-0.011	(0.014)	-0.027	(0.026)	-0.012	(0.012)	-0.030	(0.026)	-0.009	(0.009)	-0.036	(0.028)
Display	0.061	(0.061)	0.064	(0.053)	0.054	(0.065)	0.026	(0.035)	0.065	(0.063)	0.029	(0.028)	0.073	(0.063)	0.021	(0.022)	0.087	(0.067)
Feature	0.097	(0.097)	0.103	(0.085)	0.087	(0.104)	0.041	(0.056)	0.104	(0.100)	0.047	(0.045)	0.116	(0.100)	0.033	(0.036)	0.139	(0.107)

Table 3.11: Parameter estimates of the MV-MNL model for a panel data set on tuna sales (standard errors in parentheses)[a]

	Brand			Oil		Light		Volume	
	Star-Kist		CTL	Yes		Yes		Large	
Association parameters[b]									
Brand									
Star-Kist				1.548	(0.150)	2.862	(0.436)	1.937	(0.253)
CTL				-1.491	(0.203)	0.259	(0.526)	0.275	(0.327)
Oil									
Yes						-1.653	(0.755)	-1.190	(0.221)
Light									
Yes								$-\infty$[c]	–
Product-specific characteristics									
Price	-0.297	(0.014)							
Display	0.882	(0.106)							
Feature	1.508	(0.086)							

[a] Results are obtained using sufficient statistics (3.25).
[b] As the association parameters are symmetric only the upper triangular matrix is given.
[c] This combination of choices does not occur in the dataset.

Table 3.11 displays the parameter estimates and standard errors from the model specification with fixed association parameters. The parameter estimates of the marketing-mix variables are very similar to the previous specification. The advantage of this specification is that we also can interpret the association between characteristics of tuna sales. For example, given that $\hat{\psi}_{12,22} = 1.548$, it is likely that if individuals buy *Brand* Star-Kist they also choose for the *Oil*-based tuna. The opposite conclusion holds for *Brand* CTL ($\psi_{12,32} = -1.491$). Obviously, the choice for *Oil*-based tuna is negatively associated with the *Light* product. Given the large association parameter estimate for $\psi_{13,22}$ *Brand* Star-Kist apparently is market leader in low fat tuna.

To conclude, the two examples in this section show that the MV-MNL model can be used to model simultaneous multinomial decisions in a cross-sectional and in a panel context.

3.5 Conclusion

In this chapter we have introduced a novel multivariate multinomial logit specification to describe simultaneous multinomial decisions. The advantages of the new model specification over other potential model specifications are that (i) the number of parameter stays limited; (ii) there is a clear interpretation of model parameters and that; (iii) parameter estimation is feasible even if the multivariate dimension is large.

To estimate the parameters of the MV-MNL model we have proposed to use a composite likelihood function. This method limits the computational burden of a regular likelihood approach and is computationally feasible even if the multivariate dimension is large. Next, numerical problems caused by infinitely small probabilities are avoided. The resulting maximum composite likelihood estimator is consistent. A small Monte Carlo study shows that the small-sample bias of this estimator is comparable with a regular maximum likelihood estimator and that the loss in efficiency is small.

The applicability of the novel MV-MNL specification is illustrated in an application to self-reported satisfaction about life, income, social security, democracy and government. The proposed extension to panel data is illustrated using a household panel scanner data set, where we describe the purchase choice of canned tuna which we disentangle in several characteristics like brand, oil/water based and can size.

Finally, the present model specification can be extended in several directions. A possible extension is to include dynamics to the panel data model. Parameter estimation will be straightforward unless one opts for dynamics together with individual-specific effects (Honore and Kyriazidou, 2000; Carro, 2007). Other potential extensions are to adjust the model for multivariate ordered and rank ordered data or to take into account that not all choice options have to be in the consideration set of each individual.

3.A Derivations

3.A.1 Joint probabilities in MV-MNL

In this section we derive the joint probability $\Pr[Y_i = y_i]$ in the MV-MNL model taking as starting point the conditional probabilities. To derive the joint probability (from now on abbreviated as $\Pr[y]$) in the MV-MNL model, we use the identity

$$\frac{\Pr[y]}{\Pr[\mathbf{1}]} = \prod_{k=1}^{K} \frac{\Pr[y_k|y_1,\ldots,y_{k-1},1,\ldots,1]}{\Pr[y_k=1|y_1,\ldots,y_{k-1},1,\ldots,1]}. \tag{3.31}$$

which follows from the theorem of Besag (1974). The denominator in the conditional probabilities (3.2) is the same in both the numerator and denominator of (3.31) and hence drops out of the ratio. Second, the numerator of $\Pr[y_k = 1|y_1,\ldots,y_{k-1},1,\ldots,1]$ is simply proportional to 1 due to our identification restrictions. Therefore (3.31) simplifies to

$$\frac{\Pr[y]}{\Pr[\mathbf{1}]} = \prod_{k=1}^{K} \exp\Big(\alpha_{k,y_k} + X\beta_{k,y_k} + \sum_{l<k} \psi_{kl,y_k y_l} + \sum_{l>k} \psi_{kl,y_k 1}\Big). \tag{3.32}$$

Due to the restriction $\psi_{kl,y_k 1} = 0$ we obtain after rewriting

$$\frac{\Pr[y]}{\Pr[\mathbf{1}]} = \exp\Big(\sum_{k=1}^{K}\Big(\alpha_{k,y_k} + X\beta_{k,y_k} + \sum_{l>k}\psi_{kl,y_k y_l}\Big)\Big). \tag{3.33}$$

To obtain $\Pr[y]$ we use the identity

$$\Pr[y] = \frac{\Pr[y]/\Pr[\mathbf{1}]}{\sum_{s\in S}\Pr[s]/\Pr[\mathbf{1}]}, \tag{3.34}$$

where S is the set of all possible choice combinations. Substituting (3.33) in (3.34) results in

$$\Pr[y] = \frac{\exp(\mu_y)}{\sum_{s\in S}\exp(\mu_s)}, \tag{3.35}$$

where

$$\mu_y = \sum_{k=1}^{K}\left(\alpha_{k,y_k} + X\beta_{k,y_k} + \sum_{l>k}\psi_{kl,y_k y_l}\right). \tag{3.36}$$

3.A.2 Choice probability conditional on sufficient statistic

In this section we derive the panel joint choice probabilities conditional on the proposed sufficient statistics in a fixed-effects MV-MNL model of Section 3.3. If we condition on the sufficient statistic in (3.23) or (3.25), only the choice alternatives where the sufficient statistic holds are relevant, that is

$$\Pr[y_i|v_i^{(r)}] = \frac{\Pr[y_i]}{\sum_{d_i\in B_i}\Pr[d_i]}, \tag{3.37}$$

where $r = \{1,2\}$, and where B_i is the subset of alternatives which corresponds to $v_i^{(r)}$. Since we assume no dynamics we can write

$$\Pr[y_i|v_i^{(r)}] = \frac{\prod_{t=1}^{T}\Pr[y_{it}]}{\sum_{d_i\in B_i}\prod_{t=1}^{T}\Pr[d_{it}]} \tag{3.38}$$

and as the denominator of the probabilities in both the numerator and denominator are the same, this simplifies to

$$\Pr[y_i|v_i^{(r)}] = \frac{\exp(\sum_{t=1}^{T}\mu_{y_{it}})}{\sum_{d_i\in B_i}\exp(\sum_{t=1}^{T}\mu_{d_{it}})}. \tag{3.39}$$

If we opt for the sufficient statistics in (3.23), we can substitute (3.22) for $\mu_{y_{it}}$ and rewrite this as

$$\begin{aligned}\Pr[y_i|v_i^{(1)}] &= \frac{\exp(\sum_{t=1}^T \sum_{k=1}^K \left(\alpha_{ik,y_{itk}} + \sum_{l>k} \psi_{ikl,y_{itk}y_{itl}}\right))}{\sum_{d_i \in B_i} \exp(\sum_{t=1}^T \sum_{k=1}^K \left(\alpha_{ik,d_{itk}} + \sum_{l>k} \psi_{ikl,d_{itk}d_{itl}}\right))} \times \\ &\quad \frac{\exp(\sum_{t=1}^T \sum_{k=1}^K X_{it}\beta_{k,y_{itk}})}{\sum_{d_i \in B_i} \exp(\sum_{t=1}^T \sum_{k=1}^K X_{it}\beta_{k,d_{itk}})}. \end{aligned} \tag{3.40}$$

As the combination of $\alpha_{ik,j}$ and $\psi_{ikl,jh}$ is by assumption constant over time, it drops out of the equation and hence we obtain

$$\Pr[y_i|v_i^{(1)}] = \frac{\exp(\sum_{t=1}^T \sum_{k=1}^K X_{it}\beta_{k,y_{itk}})}{\sum_{d_i \in B_i} \exp(\sum_{t=1}^T \sum_{k=1}^K X_{it}\beta_{k,d_{itk}})}. \tag{3.41}$$

For the sufficient statistics in (3.25), we follow the same approach and substituting (3.22) for $\mu_{y_{it}}$ results in

$$\begin{aligned}\Pr[y_i|v_i^{(2)}] &= \frac{\exp(\sum_{t=1}^T \sum_{k=1}^K \alpha_{ik,y_{itk}})}{\sum_{d_i \in B_i} \exp(\sum_{t=1}^T \sum_{k=1}^K \alpha_{ik,d_{itk}})} \times \\ &\quad \frac{\exp(\sum_{t=1}^T \sum_{k=1}^K \left(X_{it}\beta_{k,y_{itk}} + \sum_{l>k} \psi_{kl,y_{itk}y_{itl}}\right))}{\sum_{d_i \in B_i} \exp(\sum_{t=1}^T \sum_{k=1}^K \left(X_{it}\beta_{k,d_{itk}} + \sum_{l>k} \psi_{ikl,d_{itk}d_{itl}}\right))}. \end{aligned} \tag{3.42}$$

As now only $\alpha_{ik,j}$ is constant over time, only the intercepts drop out of the equation and we obtain

$$\Pr[y_i|v_i^{(2)}] = \frac{\exp(\sum_{t=1}^T \sum_{k=1}^K X_{it} \left(\beta_{k,y_{itk}} + \sum_{l>k} \psi_{kl,y_{itk}y_{itl}}\right))}{\sum_{d_i \in B_i} \exp(\sum_{t=1}^T \sum_{k=1}^K \left(X_{it}\beta_{k,d_{itk}} + \sum_{l>k} \psi_{ikl,d_{itk}d_{itl}}\right))}. \tag{3.43}$$

3.A.3 Composite Conditional Likelihood in panel data setting

In this section we show that the composite likelihood approach is also applicable in a fixed-effects panel MV-MNL model. This section presents a panel data analog, where composite likelihood and the use of sufficient statistics is combined.

We use sufficient statistics to remove the individual-specific effects from the conditional probabilities. The sufficient statistics imply that we have to consider permutations of the choices over time. Given the panel equivalence of the specification in (3.2) any permutation over time of the choices Y_{itk}, $k = 1, \ldots, K$, yields the same set of intercepts but a different set of association parameters. Hence, we can only deal with the situation of individual-specific intercepts $\alpha_{ik,j}$ but the $\psi_{kl,jh}$ parameters have to pooled. Using sufficient statistic (3.25) we get

$$\begin{aligned}\Pr[y_{ik}|y_{il} \text{ for } l \neq k, X_i, v_{ik}^{(2)}] &= \frac{\exp(\sum_{t=1}^{T} \alpha_{ik,y_{ik}})}{\sum_{d_{ik}\in B} \exp(\sum_{t=1}^{T} \alpha_{ik,d_{ik}})} \times \\ &\quad \frac{\exp(\sum_{t=1}^{T} X_{it}\beta_{k,y_{ik}} + \sum_{l\neq k} \psi_{kl,y_{ik}y_{il}})}{\sum_{d_{ik}\in B} \exp(\sum_{t=1}^{T} X_{it}\beta_{k,d_{ik}} + \sum_{l\neq k} \psi_{kl,d_{ik}y_{il}})}. \end{aligned} \quad (3.44)$$

As the set of intercepts $\alpha_{ik,j}$ is constant over time, they drop out of the equation resulting in

$$\Pr[y_{ik}|y_{il} \text{ for } l \neq k, X_i, v_{ik}^{(2)}] = \frac{\exp(\sum_{t=1}^{T} X_{it}\beta_{k,y_{ik}} + \sum_{l\neq k} \psi_{kl,y_{ik}y_{il}})}{\sum_{d_{ik}\in B} \exp(\sum_{t=1}^{T} X_{it}\beta_{k,d_{ik}} + \sum_{l\neq k} \psi_{kl,d_{ik}y_{il}})}. \quad (3.45)$$

Hence, using the full set of conditional probabilities $\Pr[y_{ik}|y_{il} \text{ for } l \neq k, X_i, v_{ik}^{(2)}]$ in composite likelihood estimation yields an approximation of the full likelihood conditional on the sufficient statistics. As shown by the simulation study in Section 3.2.3 composite likelihood estimation in cross-sectional data finds accurate parameter estimates with only small loss of efficiency. Unreported results show that the same holds in panel data setting.

Chapter 4

Extending the Dale Model to Multivariate Ordered Responses using Composite Likelihood Estimation

4.1 Introduction

This project aims to investigate a generalization of the bivariate global odds ratio model suggested by Dale (1986) to multivariate ordered responses. Imagine a survey with several Likert (1932) scaled questions. The answers to these questions are likely to be correlated and a multivariate ordered model is useful to investigate correlations and relations to covariates. Existing estimation methods are computationally challenging or not extendable to multivariate settings. We propose an easily scalable extension of the flexible bivariate Dale model to cover for associations in multivariate ordered responses.

The Dale model is a marginal model for bivariate ordered responses. It uses global odds ratios as association measure between the responses. That is, the cumulative marginal probabilities, which can have any form, are linked via these global odds ratios. Since it can handle quite general associations between responses, this model is more flexible than other marginal models often employed. Polychoric correlation measures, such as those used in the multivariate ordered probit model, have limited flexibility since these measures assume that the same association structure underlies all combinations of response categories. That is, where a probit model yields correlations between survey questions, the Dale model gives associations between specific answers to these questions.

The flexibility of the Dale model has nonetheless not increased its popularity. To date, usage of the Dale model has been limited to two (or occasionally three) ordered responses, since parameter estimation of more general versions of the model is particularly hard. Full maximum likelihood [ML] estimation for a multivariate generalization of the Dale model has been discussed in Molenberghs and Lesaffre (1994). However, ML is computationally challenging for this class of models since there is no analytical expression for joint probabilities if more than three ordered choices are involved. Consequently, software for fitting these models are not readily available. Very few applications of the method have been reported and not much is known about the performance of the higher-dimensional model.

In this chapter we introduce a computationally feasible estimation procedure for the multivariate extension of the flexible Dale model. That is, we investigate the properties of composite likelihood estimation (see, among others, Varin et al., 2011) in estimating such a model. Composite likelihood entails replacing the full model likelihood with the likelihood of a misspecified model, which is usually arrived at by making simplifying assumptions. For example, pairwise composite likelihood approximates the full likelihood by considering all bivariate contributions to the likelihood. Higher-order contributions, such as those arising from three- or higher-way associations between responses, are assumed not to be influential in favor of reduced computational complexity. That is, interaction terms between more than two factors are discarded. In this chapter we propose to use the pairwise composite likelihood estimation method on multivariate ordered responses as an alternative for the full likelihood approach. Note that only marginal and bivariate distributions are identified. Interest thus does not lie in (fit and forecasts of) the joint distribution, but lies in marginal and bivariate distributions as well as parameter interpretation.

The outline of this chapter is as follows. We describe the bivariate Dale model as proposed by Dale (1986) in Section 4.2. Section 4.3 describes the method of composite likelihood and discusses parameter estimation. Section 4.4 studies the properties of the method through simulations. Section 4.5 applies the specification to an international survey on health care. Finally, Section 4.6 concludes the chapter.

4.2 The Bivariate Dale Model

This section describes the bivariate Dale model. Note that we fully follow Dale (1986). The contribution of this chapter lies in the multivariate extension and estimation method to be proposed in Section 4.3.

Formally, let Y_{i1} and Y_{i2} denote two ordered random variables for individual i, $i = 1, \ldots, N$, with J_1 and J_2 choice options, respectively. We specify generalized linear models for the cumulative marginal probabilities $\eta_{ik}(j_k; \alpha_k(j), \beta_k) = \Pr[Y_{ik} \leq j | X_{ik}, \alpha_k(j), \beta_k]$, $k \in \{1, 2\}$ as

$$g_k(\eta_{ik}(j; \alpha_k(j), \beta_k)) = \alpha_k(j) - X_{ik}\beta_k, \quad j_k = 1, \ldots, J_k - 1 \tag{4.1}$$

where $\alpha_k(j)$ are choice-specific threshold parameters, X_{ik} is a set of explanatory variables specific to the choice item and where β_k denotes the corresponding parameter vector. $g_k(\cdot)$ can be any link function and a typical choice is the inverse of the cumulative distribution function of the logistic distribution, which is the so-called logit function.

These cumulative marginal distributions do not incorporate potential correlations between ordered choices y_{i1} and y_{i2}. Specific to the bivariate Dale model, the association structure is specified by so called global odds ratios

$$\begin{aligned} \psi(j, h; \theta) &= \frac{\Pr[Y_{i1} \leq j, Y_{i2} \leq h | X_i, \theta] \Pr[Y_{i1} > j, Y_{i2} > h | X_i, \theta]}{\Pr[Y_{i1} > j, Y_{i2} \leq h | X_i, \theta] \Pr[Y_{i1} \leq j, Y_{i2} > h | X_i, \theta]} \\ &= \frac{F_i(j, h; \theta)[1 - \eta_{i1}(j; \theta) - \eta_{i2}(h; \theta) + F_i(j, h; \theta)]}{[\eta_{i1}(j; \theta) - F_i(j, h; \theta)][\eta_{i2}(h; \theta) - F_i(j, h; \theta)]} \end{aligned} \tag{4.2}$$

where θ resembles the model parameters, $F_i(j, h; \theta)$ denotes the joint cumulative probability $\Pr[Y_{i1} \leq j, Y_{i2} \leq h | X_i, \theta]$ and where the second line follows from standard probability theory. Note that $\psi(j, h; \theta)$ is fully symmetric, that is, $\psi(j, h; \theta) = \psi(h, j; \theta)$. It is possible to solve for $F_i(j, h; \theta)$ in (4.2) and write it in terms of $\eta_{ik}(j)$ and $\psi(j, h)$. Hence, the joint cumulative distribution for Y_{i1} and Y_{i2} is well defined, see Appendix 4.A for this Plackett (1965) distribution.

Structure can be laid on the global odds ratios $\psi(j, h; \theta)$. That is, a link function

$$h(\psi(j, h; \theta)) = \mu + \rho(j) + \kappa(h) + \omega(j, h)^1 \tag{4.3}$$

can be used and a typical choice for $h(\cdot)$ is the logarithmic function. Uniqueness constraints must be imposed on the parameters in (4.3). We choose for $\sum_{j=1}^{J_1-1} \rho(j) = \sum_{h=1}^{J_2-1} \kappa(h) = 0$ and $\sum_{j=1}^{J_1-1} \omega(j, h) = \sum_{h=1}^{J_2-1} \omega(j, h) = 0$ for all h and j, as this allows for clear interpretation. That is, parameter μ is the average (or correlation) effect, parameters $\rho(j)$ and $\kappa(h)$ are main effects over the rows and columns, respectively, and parameter $\omega(j, h)$ represents association. The set of parameters θ now consists of $\alpha_k(j)$, β_k, μ, $\rho(j)$, $\kappa(h)$ and $\omega(j, h))$. Note that $\eta_k(j)$ and $\psi(j, h)$ are functions of these parameters.

[1] Individual-specific information can be added by including covariates to this association structure, that is $h(\psi_i(j, h; \theta)) = \mu + \rho(j) + \kappa(h) + \omega(j, h) - X_{ik}\gamma(j, h)$

Then, the likelihood function needed for full maximum likelihood estimation is given by

$$\ell(\theta; Y, X) = \sum_{i=1}^{N} \ln \pi_i(y_{i1}, y_{i2}; \theta) \tag{4.4}$$

where

$$\begin{aligned} \pi_i(j, h; \theta) &= \Pr[Y_{i1} = y_{i1}, Y_{i2} = y_{i2} | X_{i1}, X_{i2}; \theta] \\ &= F_i(j, h; \theta) - F_i(j-1, h; \theta) - F_i(j, h-1; \theta) + F_i(j-1, h-1; \theta). \end{aligned} \tag{4.5}$$

The condition that (4.5) must be nonnegative is referred to as the rectangle condition (Joe, 2014) and leads to restrictions on the parameter space, see Dale (1986).

Hence, the joint probability density function of the bivariate Dale model is fully specified by the link functions and association structure. The advantage of the Dale model over other existing configurations is that the Dale model gives more information in the association structure. That is, instead of correlations between survey questions, associations between answers to these questions can be found. However, extensions to a multivariate setting are computationally challenging. Section 4.3 introduces such extension and proposes a computationally and numerically feasible estimation method.

4.3 The Multivariate Dale Model and Composite Likelihood Estimation

The model specification in the bivariate setting can easily be used to make inferences about associations and relations to covariates. However, Molenberghs and Lesaffre (1994) show that model specification and parameter estimation complicate rapidly if three or more ordered choices are investigated. That is, the expression for the full likelihood is extensive for three ordered choices and is not analytically solvable for a dimension larger than three. We propose to use composite likelihood methods with simplifying assumptions in return for computational robustness.

Composite likelihood discards the complicated full likelihood in favor of a simpler likelihood, ordinarily constructed by making simplifying assumptions such as imposing higher-order independence among response variables. A review of composite likelihood methods is given by Varin et al. (2011).

Following Varin et al. (2011), we introduce a set of events with proper density on the response vector Y. If the likelihood for each of these events are given by $\mathcal{L}_m(\theta; y)$ the composite likelihood function is

$$\mathcal{L}_C(\theta; y) = \prod_{i=1}^{N} \prod_{m=1}^{M} \mathcal{L}_m(\theta; y_i) \tag{4.6}$$

where the subscript C stands for composite likelihood. Since each component is a valid density function, the resulting composite likelihood function yields consistent estimators. Since (4.6) is a multiplication of the components, $\mathcal{L}_C(\theta; y)$ has the properties of a likelihood of a misspecified model.

Different forms of compositions have been proposed in the literature. As our purpose is to start from the bivariate Dale model, a pairwise specification is appropriate. The pairwise likelihood considers all pairs of responses, out of a total of K responses, through the specification (Cox and Reid, 2004; Varin et al., 2011)

$$\mathcal{L}_C(\theta; y) = \prod_{i=1}^{N} \prod_{k=1}^{K-1} \prod_{l=k+1}^{K} \mathcal{L}_2(y_{ik}, y_{il}; \theta), \tag{4.7}$$

where $\mathcal{L}_2(\cdot)$ is the pairwise likelihood function. Note that each component by itself is a valid density function. Only the marginal and bivariate association parameters are estimated using the pairwise composite likelihood function and higher-order associations are neglected.

Suppose that we consider K multivariate ordered responses, and want to generalize the bivariate Dale model from Section 4.2 to the multivariate situation. The pairwise log-likelihood function follows from (4.4) as

$$l_c(\theta; y, X) = \sum_{i=1}^{N} \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} \ln \pi_{ikl}(y_{ik}, y_{il}; \theta), \tag{4.8}$$

where y_{ik} denotes the realization of Y_{ik} of response k for individual i. Appropriate generalizations of (4.1) and (4.3) must also be made. Specifically, (4.1) and (4.3) are given by

$$g_k(\eta_{ik}(j; \theta)) = \alpha_k(j) - X_{ik}\beta_k, \quad j = 1, \ldots, J_k - 1 \tag{4.9}$$

$$h(\psi_{kl}(j, h; \theta)) = \mu_{kl} + \rho_{kl}(j) + \kappa_{kl}(h) + \omega_{kl}(j, h).^2 \tag{4.10}$$

Hence, where the full likelihood specification in the multivariate Dale model would involve a difficult estimation procedure using simulations (Molenberghs and Lesaffre, 1994), we introduce a relatively simple procedure using a pairwise composite likelihood method. That is, the log-likelihood only consists of a collection of bivariate Plackett distributions. Since this procedure consists of density functions which are separately valid, the estimator is consistent (Varin et al., 2011). Nonetheless, composite likelihood is an approximation of the full likelihood. The joint distribution cannot be investigated and only marginal and bivariate distributions can be investigated. Computational advan-

[2] Again, the association effects can be extended with the individual-specific effects $X_{ik}\gamma_{kl}(j, h)$.

tages outweigh this drawback if estimators are accurate and efficiency losses are small. Simulation studies in Section 4.4 show that the pairwise composite likelihood method gives accurate parameter estimates and efficiency losses are small. The method therefore is a sound alternative for full maximum likelihood; it reduces computational complexity and is appropriate for interpretation of associations in multivariate ordered data with high dimensions.

4.3.1 Parameter Inference

The set of parameters θ to be estimated consists of the marginal category thresholds $\alpha_k(j)$, the marginal effects of the covariates β_k, the association intercept parameters μ_{kl}, the association main effects $\rho_{kl}(j)$ and $\kappa_{kl}(h)$ and the association interaction effects $\omega_{kl}(j,h)$. In this section we detail parameter estimation and inference of the parameters.

For estimation purposes we use the composite likelihood approach as introduced in the previous section. That is, we maximize the composite log-likelihood function in (4.8). The analytical expression of the derivatives is given in Appendix 4.B. Note that the estimator is consistent although not efficient since the composite instead of the full likelihood is used (Varin et al., 2011). Differences between the two approaches are investigated in simulation studies in Section 4.4.

Composite likelihood assumes independence between the likelihood contributions $\mathcal{L}_2(y_{ik}, y_{il}; \theta)$ in (4.7). Therefore, standard parameter inference as in maximum likelihood does not hold. We briefly describe standard error calculation and likelihood ratio tests as described in Varin et al. (2011).

We follow an adjusted strategy to obtain parameter standard errors. The second Bartlett (1953) identity does not hold and we therefore need to distinguish between the sensitivity matrix

$$H_{\hat{\theta}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} \nabla \ell_2(\theta; y_{ik}, y_{il}) \nabla \ell_2'(\theta; y_{ik}, y_{il}) \tag{4.11}$$

and the variability matrix

$$J_{\hat{\theta}} = \frac{1}{N} \sum_{i=1}^{N} \nabla \ell_c(\theta; y_i) \nabla \ell_c'(\theta; y_i), \tag{4.12}$$

where $\nabla \ell_{\cdot}(\cdot)$ denotes the first derivative of the corresponding log-likelihood contribution (see Cox and Reid, 2004, for a broad explanation). This results in the Godambe (1960) sandwich information matrix specified as

$$G_{\hat{\theta}} = H_{\hat{\theta}} J_{\hat{\theta}}^{-1} H_{\hat{\theta}}. \tag{4.13}$$

Asymptotic results (Molenberghs and Verbeke, 2005) show that the central limit theorem holds and the estimator is asymptotically normally distributed, that is

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, G_{\theta}^{-1}). \tag{4.14}$$

Composite likelihood therefore yields consistent estimators. Efficiency is nonetheless lost, which can be investigated by comparing the Fisher information matrix from maximum likelihood estimation with the Godambe (1960) sandwich information matrix from our estimation methodology. Hence, we can investigate the influence of neglecting higher order associations in our multivariate ordered choice model, given that we know the full likelihood specification.

In line with the model specification it would be of interest to test for joint significance of the association parameters in $\psi_{kl}(j, h; \theta)$. Several adjusted likelihood ratio [LR] tests have been proposed and we opt for the method of Kent (1982) including a Satterthwaite (1946) adjustment. That is, the LR test-statistic is defined as

$$LR = \frac{2\nu}{Q\bar{\lambda}}(\ell_c(\theta; y) - \ell_c(\theta_R; y)), \tag{4.15}$$

where Q is the number of restricted parameters,

$$\nu = \frac{(\sum_{q=1}^{Q} \lambda_q)^2}{\sum_{q=1}^{Q} \lambda_q^2} \tag{4.16}$$

and where $\lambda_1, \ldots, \lambda_Q$ denote the eigenvalues of $(G_R(H^{-1})_R)^{-1}$ with G_R the $Q \times Q$ sub-matrix of the Godambe (1960) sandwich information matrix corresponding to the restricted parameters under the null hypothesis. Moreover, $\bar{\lambda}$ denotes the average of the eigenvalues. Then, (4.15) has a χ^2-distribution with ν degrees of freedom under the null hypothesis. Hence, the LR-test compares the composite log likelihood value with the restricted likelihood value as standard likelihood ratio tests do. The distribution is nonetheless shifted taking into account the assumption of independence of likelihood contributions in (4.7).

Model comparison can easily be done using adjusted information criteria. Gao and Song (2010) derive the composite likelihood analogue of the Bayesian information criterion [BIC], that is

$$BIC = -2\ell_c(\theta; y) + dim(\theta) \ln n \tag{4.17}$$

where $dim(\theta)$ replaces the number of parameters in standard BIC. This effective number of parameters equals $tr(H_{\theta}G_{\theta}^{-1})$ where $tr(\cdot)$ is the trace of the matrix. This criterion can then be used to compare model specifications with different sets of covariates in the multivariate ordered choice model.

4.4 Simulation Study

In this section we conduct a Monte Carlo study to investigate the properties of the pairwise composite likelihood method in comparison to the full likelihood approach. We especially focus on potential small-sample biases and efficiency losses due to the misspecified likelihood. For comparison reasons, we add the multivariate ordered probit specification to our analysis. We again use a pairwise composite likelihood method following Bhat et al. (2010).

As usual, we simulate the data from the joint distribution, that is the multivariate Dale model following Molenberghs and Lesaffre (1994). This specification does not have an analytical expression when the dimension is larger than 3 (Molenberghs and Lesaffre, 1994). We therefore do a simulation study where we have 3 ordered choices. We then also are able to use full maximum likelihood as benchmark method. Note that the higher-order associations - disregarded in pairwise composite likelihood - now consist of the three-way global odds ratios.

We consider two data generating processes [DGPs]. First, we investigate what the multivariate Dale model can find more than multivariate probit. That is, where the correlation effects in this DGP are close to 0 we add positive and negative association effects for the specific combinations of choices. Second, we investigate the impact of neglecting higher order associations which actually have significant effect on the multivariate choices. Hence, these DGPs are extreme cases to show what information might be lost in practical applications.

We use logit link functions for the marginal effect $\eta_k(j)$ and a log transformation for $\psi_{kl}(j,h)$ since positive global odds ratios are obliged. The number of choice options J_k per ordered choice is 3. We consider a likely sample size of $N = 1000$ observations. As covariate X_{ik} we take a normally distributed variable which is different over all ordered choices. The effect of the covariate is positive for the first ordered choice, non-existent for the second and negative for the third. We further take a simplified version of $h(\psi_{ikl}(j,h;\theta))$ where $\rho_{kl}(j) = \kappa_{kl}(h) = 0$. See Tables 4.1 and 4.2 for the values of our DGP-parameters.

Table 4.1: Mean and RMSE of the parameter estimates for the multivariate Dale model parameters based on the first simulation study[a] with $N = 1000$ (1000 replications)

		full likelihood		composite likelihood		ordered probit	
	θ	$\hat{\theta}$	RMSE	$\hat{\theta}$	RMSE	$\hat{\theta}$	RMSE
$\alpha_1(1)$	-0.8	-0.804	0.070	-0.804	0.070	-0.485	0.042
$\alpha_1(2)$	1	0.998	0.076	0.998	0.076	0.602	0.045
$\alpha_2(1)$	-0.45	-0.450	0.064	-0.450	0.064	-0.280	0.040
$\alpha_2(2)$	0.85	0.851	0.068	0.851	0.068	0.527	0.041
$\alpha_3(1)$	-0.95	-0.955	0.077	-0.955	0.077	-0.575	0.046
$\alpha_3(2)$	0.95	0.948	0.073	0.948	0.073	0.571	0.043
β_1	0.85	0.853	0.064	0.853	0.064	0.510	0.038
β_2	0	0.000	0.060	0.000	0.060	0.000	0.037
β_3	-0.85	-0.856	0.069	-0.855	0.069	-0.510	0.040
$\mu_{1,2}$	0.1	0.101	0.114	0.101	0.114		
$\mu_{1,3}$	0.1	0.103	0.113	0.104	0.113		
$\mu_{2,3}$	0.05	0.043	0.114	0.042	0.114		
$\omega_{1,2}(2,1)$	0.2	0.199	0.112	0.199	0.111		
$\omega_{1,2}(1,2)$	-0.3	-0.299	0.104	-0.298	0.103		
$\omega_{1,2}(2,2)$	-0.4	-0.401	0.118	-0.401	0.117		
$\omega_{1,3}(2,1)$	-0.2	-0.198	0.119	-0.197	0.119		
$\omega_{1,3}(1,2)$	0.1	0.100	0.122	0.099	0.121		
$\omega_{1,3}(2,2)$	0.45	0.455	0.123	0.455	0.121		
$\omega_{2,3}(2,1)$	0.05	0.048	0.110	0.048	0.109		
$\omega_{2,3}(1,2)$	-0.05	-0.043	0.105	-0.043	0.105		
$\omega_{2,3}(2,2)$	-0.25	-0.263	0.114	-0.262	0.114		
$\psi_{1,2,3}(1,1,1)$	0.95	1.019	0.352				
$\psi_{1,2,3}(2,1,1)$	0.95	1.011	0.367				
$\psi_{1,2,3}(1,2,1)$	1.25	1.364	0.561				
$\psi_{1,2,3}(2,2,1)$	1.2	1.290	0.525				
$\psi_{1,2,3}(1,1,2)$	0.85	0.903	0.314				
$\psi_{1,2,3}(2,1,2)$	0.8	0.844	0.308				
$\psi_{1,2,3}(1,2,2)$	1	1.083	0.438				
$\psi_{1,2,3}(2,2,2)$	1.05	1.138	0.494				
$\sigma_{1,2}$[b]						0.051	0.042
$\sigma_{1,3}$						0.036	0.041
$\sigma_{2,3}$						0.020	0.041

[a] The first simulation study investigates the advantage over multivariate ordinal probit by having insignificant correlation effects.
[b] σ denotes the correlation between the choice items in multivariate ordinal probit

Table 4.1 shows the mean and root mean squared error [RMSE] of the estimates from all three methods over 1000 replications for the first DGP. It is clear that the estimation methods result in small small-sample biases. Most importantly, the estimates of our composite likelihood approach are highly comparable to the full likelihood estimates. The largest difference is only 0.0009.

To analyze the loss in efficiency between the composite likelihood and full likelihood approach we compare the RMSEs of both methods. These measures are highly comparable. At a maximum, the loss in efficiency is only 0.3 percent. Since we find comparable sample biases and small efficiency losses, the advantages of computation time and scalability are in favor of our proposed estimation method.

The parameter estimates of the multivariate ordered probit model are shown in the final columns of Table 4.1. From the insignificant average estimates of the correlation parameters we expect no relation between the ordered choices. However, different conclusions can be drawn from the multivariate Dale model. 4 out of 9 association effects significantly differ from 0. Since some are positive and some are negative, these effects are vanished in the multivariate ordered probit specification. The multivariate Dale model does capture this association structure and thus gives more information than standard polychoric correlations.

Table 4.2: Mean and RMSE of the parameter estimates for the multivariate Dale model parameters based on the second simulation study[a] with $N = 1000$ (1000 replications)

		full likelihood		composite likelihood		ordered probit	
	θ	$\hat{\theta}$	RMSE	$\hat{\theta}$	RMSE	$\hat{\theta}$	RMSE
$\alpha_1(1)$	-0.8	-0.799	0.089	-0.799	0.069	-0.504	0.042
$\alpha_1(2)$	-0.05	-0.051	0.063	-0.051	0.063	-0.043	0.040
$\alpha_2(1)$	-1.05	-1.050	0.104	-1.050	0.074	-0.650	0.044
$\alpha_2(2)$	-0.25	-0.249	0.068	-0.249	0.066	-0.163	0.041
$\alpha_3(1)$	-1.05	-1.050	0.104	-1.051	0.073	-0.650	0.043
$\alpha_3(2)$	-0.25	-0.249	0.067	-0.249	0.065	-0.163	0.040
β_1	0.1	0.102	0.051	0.102	0.053	0.063	0.034
β_2	0	0.001	0.046	0.001	0.051	0.001	0.032
β_3	-0.1	-0.100	0.047	-0.099	0.050	-0.061	0.032
$\mu_{1,2}$	1.4	1.410	0.156	1.409	0.121		
$\mu_{1,3}$	1.4	1.412	0.157	1.411	0.122		
$\mu_{2,3}$	1.9	1.910	0.186	1.908	0.129		
$\omega_{1,2}(2,1)$	0.1	0.105	0.092	0.104	0.092		
$\omega_{1,2}(1,2)$	-0.3	-0.300	0.083	-0.300	0.080		
$\omega_{1,2}(2,2)$	0.3	0.299	0.082	0.299	0.080		
$\omega_{1,3}(2,1)$	0.1	0.103	0.092	0.102	0.092		
$\omega_{1,3}(1,2)$	-0.3	-0.300	0.082	-0.299	0.079		
$\omega_{1,3}(2,2)$	0.3	0.299	0.084	0.299	0.081		
$\omega_{2,3}(2,1)$	-0.05	-0.044	0.098	-0.044	0.098		
$\omega_{2,3}(1,2)$	-0.05	-0.050	0.097	-0.050	0.098		
$\omega_{2,3}(2,2)$	0.3	0.296	0.094	0.296	0.092		
$\psi_{1,2,3}(1,1,1)$	3.25	3.512	1.321				
$\psi_{1,2,3}(2,1,1)$	1	1.096	0.438				
$\psi_{1,2,3}(1,2,1)$	1.75	1.895	0.735				
$\psi_{1,2,3}(2,2,1)$	0.5	0.540	0.204				
$\psi_{1,2,3}(1,1,2)$	1.75	1.866	0.704				
$\psi_{1,2,3}(2,1,2)$	0.5	0.536	0.205				
$\psi_{1,2,3}(1,2,2)$	1	1.053	0.356				
$\psi_{1,2,3}(2,2,2)$	0.25	0.260	0.087				
$\sigma_{1,2}$[b]						0.508	0.035
$\sigma_{1,3}$						0.508	0.035
$\sigma_{2,3}$						0.638	0.029

[a] The second simulation study investigates the harm of neglecting three-way global cross ratios which have significant effect on the multivariate choices.

[b] σ denotes the correlation between the choice items in multivariate ordinal probit

Table 4.2 shows the results of the second simulation study, where higher order associations in the DGP have a significant effect on multivariate choice. Again, differences of small-sample biases are small and negligible. The loss in efficiency is somewhat larger, since significant effects are neglected in the pairwise composite likelihood approach. The largest efficiency loss found in this simulation study is 10 percent. It depends on the researcher whether (s)he thinks that this loss in efficiency is acceptable. Since parameter estimates are highly comparable, the harm of neglecting higher order associations is only small.

As a final advantage of the composite likelihood approach over the other approaches we compare computation times. Note that other sets of parameters are estimated and the comparison is not entirely fair. Nonetheless, differences are interesting since they are large. Composite likelihood is on average 9 times faster than multivariate probit and more than 16 times faster than the maximum likelihood approach. This holds for a dimension of 3 but the advantage will only increase if the dimension gets larger. Namely, the number of parameters and the complexity increases for the full likelihood approach, and multivariate probit needs more and more integrals to be investigated.

In sum, the simulation study shows that our composite likelihood approach has the advantage of being (i) comparable to the full likelihood approach in small-sample bias and efficiency; (ii) able to find associations between specific choice options and; (iii) reasonably fast in computation. Together with its scalability, we think this to be reasons to favor our estimation method over other available methods for parameter interpretation purposes. We therefore empirically illustrate the estimation method in Section 4.5.

4.5 Application

In this section we apply the multivariate Dale model to International Social Survey Program [ISSP] data from 2011.[3] The data set consists of 33900 respondents from 30 countries. We choose 8 questions on satisfaction on the health system. These questions are on Likert scale basis and the rating scales are $J_1 = 4$, $J_2 = \cdots = J_7 = 5$ and $J_8 = 7$. The questions and scales are shown in Table 4.C.1. As explanatory variables for the ordered choices we use dummy variables for age lower than 35 and age higher than 64 (compared to age 35-64), male (compared to female), employed in the past and never employed (compared to employed respondents) and country-specific dummies (compared to the US). Although possible, we do not add covariates in the link function for associations, but leave the flexibility of all correlation (μ), main (ρ and κ) and association (ω) effects. As the survey questions are all health-related and asked to the same respondents, the answers are likely to be correlated.

[3] This data is openly available at http://www.issp.org/page.php?pageId=4

The model specification introduced in this chapter is therefore highly applicable, especially if one is interested in interpretation of the association structure.

We use the composite likelihood method from Section 4.3 since there is no analytical expression for the full likelihood specification with 8 correlated ordered choices. The composite likelihood method highly simplifies the analysis as it neglects three- to eight-way associations. Since there is no analytical expression for the joint distribution, focus lies on parameter interpretation and marginal and bivariate distributions. The simulation study in Section 4.4 showed that neglecting higher order associations does not tremendously harm small-sample bias and efficiency of the estimator.

We first test for the necessity of the parameters in the global odds ratios in (4.10). The LR-statistic in (4.15) of only correlation effects tested against univariate analyzes equals 43188.0 (17.5 degrees of freedom). This statistic clearly shows that independence is rejected. Testing for the addition of main effects yields an LR-statistic of 5403.0 (123.5 degrees of freedom). Finally, testing for the addition of association effects yields an LR-statistic of 15436.3 (217.4 degrees of freedom). These conclusions clearly are in favor of a multivariate analysis opposed to separate univariate analyses.

Table 4.3 displays the intercept estimates, which represent the baseline respondent (an employed male of age 35 to 64 from the United States). On average, this respondent disagrees that the health care system will improve, disagrees that the government should only provide limited services and in the end is fairly satisfied about the health care system.

Table 4.C.2 shows the parameter estimates belonging to the covariates. Unreported results show that the difference of these estimates to a univariate specification or only small (0.06 at a maximum). That is, the relation between covariates and the ordered dependent variables is not influenced much by taking associations into account. From the table we see that respondents of age higher than 64 ceteris paribus more often agree with the proposition. That is, they are more satisfied about the health care system and demand less for changes. The same relations hold for respondents who have never been employed. They further on average highly believe that improvements will come in the future. Respondents of age lower than 35 on average do not think they will get the best treatment available by the doctor of their choice. Finally, some differences between countries can be given. All statements are compared to the base country, the United States. Respondents from Great Britain are rather negative about the health care system, while respondents from Sweden are overly positive. The countries which are most and least satisfied about the health care system are Sweden and Bulgaria, respectively.

Table 4.3: Intercept estimates for the pairwise composite likelihood method in the ISSP health care survey (standard errors in parentheses)

Question	1		2		3		4		5		6		7		8	
α_1	-3.234	(0.054)	-3.716	(0.059)	-2.032	(0.054)	-3.155	(0.056)	-2.219	(0.052)	-2.325	(0.053)	-1.475	(0.052)	-3.835	(0.057)
α_2	-0.143	(0.049)	-1.562	(0.052)	0.098	(0.050)	-1.531	(0.051)	-0.349	(0.047)	-0.176	(0.050)	0.446	(0.051)	-1.927	(0.051)
α_3	2.009	(0.051)	-0.309	(0.048)	1.197	(0.050)	-0.701	(0.050)	0.902	(0.046)	1.254	(0.050)	1.750	(0.050)	0.194	(0.049)
α_4	–		1.552	(0.050)	3.150	(0.052)	1.026	(0.046)	2.744	(0.048)	2.906	(0.050)	3.431	(0.049)	1.243	(0.046)
α_5	–		–		–		–		–		–		–		2.409	(0.048)
α_6	–		–		–		–		–		–		–		3.596	(0.053)

Table 4.4: Correlation effect estimates for the pairwise composite likelihood method in the ISSP health care survey (standard errors in parentheses)

Question	1		2		3		4		5		6		7	
1														
2	0.893	(0.023)												
3	0.196	(0.022)	0.564	(0.020)										
4	0.232	(0.021)	0.458	(0.022)	0.912	(0.025)								
5	-0.964	(0.023)	-0.585	(0.019)	0.028	(0.019)	0.270	(0.021)						
6	0.752	(0.027)	0.794	(0.022)	0.363	(0.020)	0.298	(0.022)	-0.659	(0.023)				
7	0.554	(0.025)	0.633	(0.020)	0.303	(0.020)	0.323	(0.021)	-0.406	(0.020)	2.003	(0.028)		
8	1.650	(0.040)	1.160	(0.024)	0.425	(0.022)	0.362	(0.025)	-1.029	(0.027)	1.437	(0.030)	1.081	(0.025)

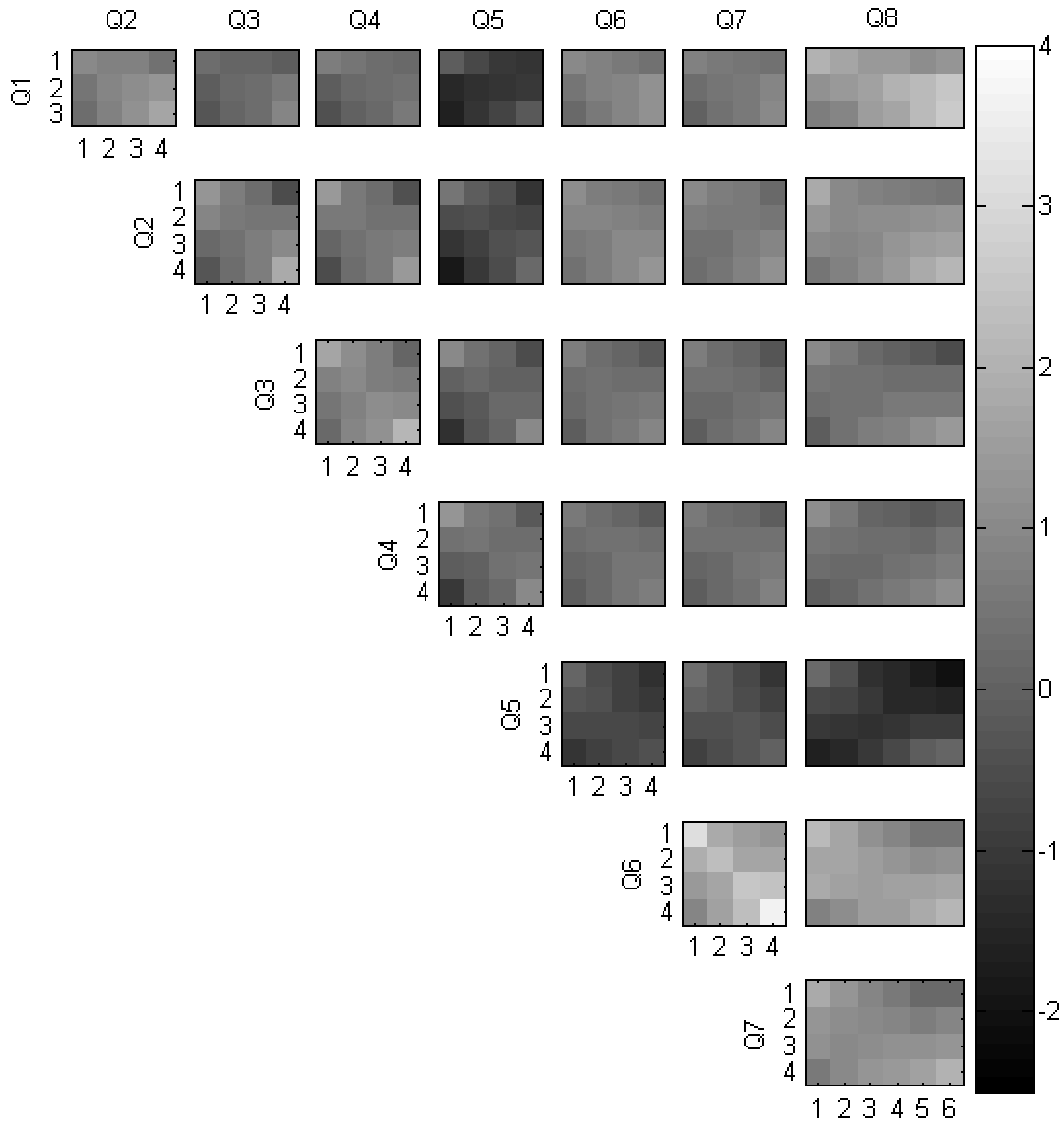
More interest goes to the special feature of our model specification. The multivariate Dale model gives associations which indicate what correlation structure there is in the answers of respondents. First, the correlation effects represented by μ are shown in Table 4.4. There especially is high positive association between the questions on best treatment (Q6) and doctor of own choice (Q7). Further, respondents satisfied with the health care system (Q8) seem also to expect future improvements (Q2) and respondents in favor of limited services (Q4) also tend to think that health care is too excessively used (Q3).

Figure 4.1 displays log global odds ratios for all pairs of ordered choices.[4] The subfigure at grid $(3, 5)$ clearly shows that associations might differ over combinations of answers. Where $\mu_{3,5}$ indicates that correlation between answers to inefficiency of the system and excessive use does not significantly differ from 0, we clearly see associations between specific answers. That is, respondents with extreme answers, answer extreme to both questions. This effect is significant. Nonetheless, since some associations are negative where others are positive, $\mu_{3,5}$ does not capture this attribute of the data. The sub-figures at grids $(1, 8)$ and $(3, 4)$ show other pairs of questions and also show large and significant differences between combinations of answers. For example, grid $(1, 8)$ shows the association in the answers to questions on limited health care services and the inefficiency of the system. Clearly, main effects $\rho(i)$ for the limited services play a role. Further, $\kappa(j)$ for health care satisfaction adds information to the analysis with the question on the need of changes (grid $(3, 4)$).

In sum, the multivariate Dale model can be used to analyze correlated answers in this ISSP survey. Interpretations of effects of covariates are clear and the LR-test clearly shows that the answers to the survey questions are correlated. If the associations are especially of interest, the multivariate Dale model is favored over univariate analyses. Only reporting correlations might lead to misleading conclusions since some answers to the questions might be more associated with one another than others. The main and associations effects clearly give extra information. Therefore, together with the fact that there is no analytical expression for the full likelihood specification, analysis using a multivariate Dale model is favorable over multivariate ordered probit and the full maximum likelihood approach.

[4]For the reason of space, we do not report all results. All parameter estimates including standard errors are available upon request.

Figure 4.1: Log global cross ratios for the pairwise composite likelihood method in the ISSP health care survey



4.6 Conclusion

In this project we have investigated the multivariate extension of the bivariate Dale (1986) model. This model is highly appropriate to analyze correlated ordered responses. We have advocated the pairwise composite likelihood method to overcome computational and analytical problems. Where the full likelihood function has no analytical expression if the number of ordered choices is larger than three, the composite likelihood method circumvents this problem. Opposed to available methods in the literature, the computational burden is only small.

Although we use a misspecified likelihood function, simulation studies show that it yields comparable and negligible small-sample biases. Further, efficiency losses are small. A comparison with the multivariate ordered probit model shows that more information is found due to the association structure in the global odds ratios.

In an application, we have used the composite likelihood method on ISSP health care survey data and found plausible results. Note that the full likelihood approach could not have been used analytically given the high dimension of 8 ordered responses.

In sum, the multivariate Dale model together with the proposed estimation method is highly applicable and useful in empirical research when interest lies in parameter interpretation. This research gives the opportunity to estimate relations between many responses in a reasonable amount of time with high accuracy.

4.A Plackett Distribution

It can be shown by rewriting that the joint cumulative probability $F_i(j,h;\theta)$ of the bivariate Dale model can be expressed as (see Dale, 1986)

$$F_i(j,h;\theta) = \begin{cases} \frac{1}{2}(\psi_i(j,h)-1)^{-1}[1+(\eta_{i1}(j)+\eta_{i2}(h))(\psi_i(j,h)-1)- \\ \quad S(\eta_{i1}(j),\eta_{i2}(h),\psi_i(j,h))] \quad \text{if } \psi_i(j,h)\neq 1; \\ \eta_{i1}(j)\eta_{i2}(h) \quad \text{otherwise} \end{cases} \tag{4.18}$$

where

$$S(\eta_{i1}(j),\eta_{i2}(h),\psi_i(j,h)) = \sqrt{\begin{array}{l}[1+(\eta_{i1}(j)+\eta_{i2}(h))(\psi_i(j,h)-1)]^2+ \\ 4\psi_i(j,h)(1-\psi_i(j,h))\eta_{i1}(j)\eta_{i2}(h).\end{array}} \tag{4.19}$$

4.B Derivatives

We use a quasi-Newton method for parameter estimation. It is therefore helpful to have the first-order derivatives of the log-likelihood with respect to the model parameters available to improve convergence. Let

$$\begin{aligned} l_{ikl}(j,h) &= \log \pi_{ikl}(j,h;\theta) \\ l_{ikl} &= \sum_{j=1}^{J_k}\sum_{h=1}^{J_l} l_{ikl}(j,h). \end{aligned} \tag{4.20}$$

The parameter vector θ contains five types of parameters, namely those of the form θ_k, $\theta_k(j)$, θ_{kl}, $\theta_{kl}(j)$ and $\theta_{kl}(j,h)$. The different derivatives for these types of parameters are

$$\begin{aligned} \frac{\partial l_c}{\partial \theta_k} &= \sum_{i=1}^{N}\sum_{l\neq k}\sum_{j=1}^{J_k}\sum_{h=1}^{J_l} \frac{\partial l_{ikl}}{\partial F_{ikl}(j,h)}\cdot\frac{\partial F_{ikl}(j,h)}{\partial \eta_{ik}(j)}\cdot\frac{\partial \eta_{ik}(j)}{\partial \theta_k} \\ \frac{\partial l_c}{\partial \theta_k(j)} &= \sum_{i=1}^{N}\sum_{l\neq k}\sum_{h=1}^{J_l} \frac{\partial l_{ikl}}{\partial F_{ikl}(j,h)}\cdot\frac{\partial F_{ikl}(j,h)}{\partial \eta_{ik}(j)}\cdot\frac{\partial \eta_{ik}(j)}{\partial \theta_k(j)} \\ \frac{\partial l_c}{\partial \theta_{kl}} &= \sum_{i=1}^{N}\sum_{j=1}^{J_k}\sum_{h=1}^{J_l} \frac{\partial l_{ikl}}{\partial F_{ikl}(j,h)}\cdot\frac{\partial F_{ikl}(j,h)}{\partial \psi_{ikl}(j,h)}\cdot\frac{\partial \psi_{ikl}(j,h)}{\partial \theta_{kl}} \\ \frac{\partial l_c}{\partial \theta_{kl}(j)} &= \sum_{i=1}^{N}\sum_{j=1}^{J_l} \frac{\partial l_{ikl}}{\partial F_{ikl}(j,h)}\cdot\frac{\partial F_{ikl}(j,h)}{\partial \psi_{ikl}(j,h)}\cdot\frac{\partial \psi_{ikl}(j,h)}{\partial \theta_{kl}(j)} \\ \frac{\partial l_c}{\partial \theta_{kl}(j,h)} &= \sum_{i=1}^{N} \frac{\partial l_{ikl}}{\partial F_{ikl}(j,h)}\cdot\frac{\partial F_{ikl}(j,h)}{\partial \psi_{ikl}(j,h)}\cdot\frac{\partial \psi_{ikl}(j,h)}{\partial \theta_{kl}(j,h)}. \end{aligned} \tag{4.21}$$

The derivatives of l_{ikl} with respect to $F_{ikl}(j,h)$ are common to all of (4.21) and write

$$\begin{aligned} \frac{\partial l_{ikl}}{\partial F_{ikl}(j,h)} &= \zeta_{ikl}(j,h) - \zeta_{ikl}(j+1,h) - \zeta_{ikl}(j,h+1) + \zeta_{ikl}(j+1,h+1) \\ \frac{\partial l_{ikl}}{\partial F_{ikl}(J_k,h)} &= \zeta_{ikl}(J_k,h) - \zeta_{ikl}(J_k,h+1) \\ \frac{\partial l_{ikl}}{\partial F_{ikl}(j,J_l)} &= \zeta_{ikl}(j,J_l) - \zeta_{ikl}(j+1,J_l) \\ \frac{\partial l_{ikl}}{\partial F_{ikl}(J_k,J_l)} &= \zeta_{ikl}(J_k,J_l) \end{aligned} \tag{4.22}$$

where $\zeta_{ikl}(j,h) = 1/\pi_{ikl}(j,h)$.

Omitting indices i, j and h for convenience, the partial derivatives of F_{kl} with respect to η_k and ψ_{kl} are given by

$$\frac{\partial F_{kl}}{\partial \eta_k} = \begin{cases} \eta_l & \text{if } \psi_{kl} = 1; \\ \frac{1}{2}(1 - S^{-1}(\eta_k, \eta_l, \psi_{kl})(1 + \psi_{kl}(\eta_k - \eta_l) - \eta_k - \eta_l)) & \text{otherwise} \end{cases} \tag{4.23}$$

$$\frac{\partial F_{kl}}{\partial \psi_{kl}} = \begin{cases} 0 & \text{if } \psi_{kl} = 1; \\ \frac{1}{2}[1 - S(\eta_k, \eta_l, \psi_{kl}) + (\psi_{kl} - 1)(\eta_k + \eta_l - 2\eta_k\eta_l)] \times & \\ \quad S^{-1}(\eta_k, \eta_l, \psi_{kl})(\psi_{kl} - 1)^{-2} & \text{otherwise.} \end{cases} \tag{4.24}$$

The derivatives of η_k and ψ_{kl} with respect to the parameters depend on the choice of the link functions $g_k(\cdot)$ and $h_{kl}(\cdot)$. In case that the logarithmic link function and the restrictions (4.3) are used for the association parameters we have that

$$\begin{aligned} \frac{\partial \psi_{ikl}(j,h)}{\partial \mu_{kl}} &= \psi_{ikl}(j,h) \\ \frac{\partial \psi_{ikl}(j,h)}{\partial \rho_{kl}(j)} &= \psi_{ikl}(j,h) - \psi_{ikl}(1,h) \\ \frac{\partial \psi_{ikl}(j,h)}{\partial \kappa_{kl}(h)} &= \psi_{ikl}(j,h) - \psi_{ikl}(j,1) \\ \frac{\partial \psi_{ikl}(j,h)}{\partial \omega_{kl}(j,h)} &= \psi_{ikl}(j,h) - \psi_{ikl}(j,1) - \psi_{ikl}(1,h) + \psi_{ikl}(1,1). \end{aligned} \tag{4.25}$$

The derivatives of η_k with respect to the marginal parameters under the logit function are given by

$$\frac{\partial \eta_{ik}(j)}{\partial \alpha_k(j)} = \eta_{ik}(j)(1 - \eta_{ik}(j)) \tag{4.26}$$

$$\frac{\partial \eta_{ik}(j)}{\partial \beta_k} = -\eta_{ik}(j)(1 - \eta_{ik}(j))X_{ik}. \tag{4.27}$$

4.C Tables

Table 4.C.1: List of ordered choices including Likert scale answers to the ISSP health care survey

	1	2	3	4	5	6	7
Q1	**In general, would you say that the health care system in [country]**						
	Needs no change	Needs a few changes	Needs many changes	Needs to be completely changed	–	–	–
Q2	**In the next few years the health care system in [country] will improve.**						
	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	–	–
Q3	**People use health care services more than necessary.**						
	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	–	–
Q4	**The government should provide only limited health care services.**						
	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	–	–
Q5	**In general, the health care system in [country] is inefficient.**						
	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	–	–
Q6	**How likely is it that if you become seriously ill, you would get the best treatment available in [country]?**						
	It's certain I would get	It's likely I would get	Equal change of getting or net getting	It's likely I would not get	It's certain I would not get	–	–
Q7	**How likely is it that if you become seriously ill, you would get treatment from the doctor of your choice?**						
	It's certain I would get	It's likely I would get	Equal change of getting or net getting	It's likely I would not get	It's certain I would not get	–	–
Q8	**In general, how satisfied are you with the health care system in [country]?**						
	Completely satisfied	Very satisfied	Fairly satisfied	Neither satisfied nor dissatisfied	Fairly dissatisfied	Very dissatisfied	Completely dissatisfied

Table 4.C.2: Parameter estimates for covariates for the pairwise composite likelihood method in the ISSP health care survey (standard errors in parentheses)

Question	1		2		3		4		5		6		7		8	
age lower than 35	-0.168	(0.045)	-0.140	(0.021)	-0.013	(0.021)	0.052	(0.021)	0.075	(0.021)	0.101	(0.021)	0.121	(0.020)	0.038	(0.021)
age higher than 64	-0.357	(0.043)	-0.278	(0.028)	-0.119	(0.028)	-0.215	(0.028)	0.196	(0.027)	-0.323	(0.028)	-0.321	(0.027)	-0.460	(0.028)
male	-0.141	(0.035)	-0.139	(0.018)	-0.219	(0.018)	-0.176	(0.018)	0.054	(0.018)	-0.031	(0.018)	-0.009	(0.018)	-0.115	(0.018)
employed in the past	-0.013	(0.036)	-0.094	(0.022)	0.151	(0.022)	0.064	(0.022)	0.015	(0.022)	0.086	(0.022)	0.062	(0.022)	0.075	(0.022)
never employed	-0.317	(0.040)	-0.516	(0.031)	0.066	(0.032)	-0.289	(0.032)	0.150	(0.031)	-0.134	(0.031)	-0.195	(0.031)	-0.315	(0.030)
Australia	0.156	(0.060)	-0.339	(0.061)	0.230	(0.058)	0.290	(0.058)	-0.181	(0.054)	-0.561	(0.055)	0.285	(0.054)	-0.282	(0.056)
Belgium - Flanders	-0.981	(0.080)	-0.503	(0.079)	0.178	(0.078)	0.323	(0.075)	0.583	(0.073)	-0.394	(0.074)	0.275	(0.074)	-0.853	(0.075)
Belgium - Wallonia	-1.441	(0.077)	-0.894	(0.067)	-0.189	(0.068)	0.554	(0.065)	1.643	(0.065)	-0.936	(0.072)	-0.385	(0.071)	-1.250	(0.070)
Bulgaria	1.873	(0.077)	-0.281	(0.075)	1.458	(0.076)	-0.086	(0.079)	-1.741	(0.078)	1.009	(0.076)	1.124	(0.075)	1.530	(0.071)
Switzerland	-0.204	(0.073)	-1.635	(0.072)	0.046	(0.070)	-0.153	(0.069)	-0.055	(0.067)	-1.200	(0.070)	-0.503	(0.070)	-0.640	(0.067)
Chili	1.699	(0.058)	-0.903	(0.057)	0.991	(0.058)	0.210	(0.056)	-0.801	(0.054)	0.977	(0.053)	1.657	(0.053)	1.356	(0.052)
Czech Republic	-0.017	(0.065)	-0.745	(0.061)	0.331	(0.061)	0.339	(0.060)	-0.020	(0.060)	0.020	(0.062)	0.689	(0.062)	0.087	(0.063)
Germany - East	-0.838	(0.111)	-0.606	(0.103)	0.250	(0.098)	0.291	(0.098)	0.228	(0.097)	-0.262	(0.099)	0.616	(0.098)	-0.517	(0.094)
Germany - West	-0.827	(0.084)	-0.583	(0.077)	-0.116	(0.079)	0.489	(0.077)	0.900	(0.074)	-0.612	(0.079)	0.330	(0.078)	-0.919	(0.078)
Denmark	-0.769	(0.085)	-1.191	(0.071)	-0.132	(0.071)	0.970	(0.067)	0.994	(0.065)	-1.193	(0.076)	0.757	(0.071)	-1.233	(0.073)
Finland	-0.574	(0.083)	0.148	(0.079)	0.659	(0.075)	0.089	(0.070)	0.566	(0.072)	-0.709	(0.076)	1.092	(0.075)	-0.680	(0.076)
France	-0.535	(0.058)	0.244	(0.054)	-0.782	(0.054)	0.759	(0.053)	1.731	(0.053)	-0.679	(0.056)	0.088	(0.055)	-0.563	(0.056)
Great Britain	2.031	(0.079)	0.404	(0.076)	0.766	(0.073)	0.789	(0.075)	-0.608	(0.069)	0.195	(0.072)	0.624	(0.071)	1.224	(0.073)
Croatia	0.507	(0.078)	-0.507	(0.076)	0.917	(0.074)	0.175	(0.076)	0.239	(0.075)	0.335	(0.073)	0.724	(0.074)	0.277	(0.073)
Israel	0.835	(0.073)	-1.231	(0.076)	0.952	(0.076)	0.192	(0.070)	0.958	(0.079)	-0.543	(0.075)	0.480	(0.074)	-0.331	(0.073)
Japan	-0.500	(0.083)	0.204	(0.072)	-0.157	(0.074)	-0.116	(0.071)	-0.681	(0.073)	1.250	(0.078)	1.085	(0.080)	0.561	(0.077)
South-Korea	-0.383	(0.067)	-2.288	(0.068)	-0.409	(0.064)	-0.643	(0.062)	-0.156	(0.063)	-0.522	(0.062)	0.200	(0.062)	-0.091	(0.063)
Lithuania	0.464	(0.079)	-0.321	(0.080)	1.575	(0.081)	-0.147	(0.082)	-0.095	(0.080)	0.444	(0.074)	0.614	(0.079)	0.480	(0.079)
The Netherlands	-0.464	(0.067)	-0.180	(0.065)	-0.058	(0.067)	0.031	(0.067)	0.042	(0.065)	-0.767	(0.066)	0.360	(0.066)	-0.906	(0.064)
Norway	0.487	(0.076)	-0.784	(0.072)	0.255	(0.071)	0.548	(0.068)	-0.061	(0.068)	-0.781	(0.071)	1.551	(0.068)	-0.648	(0.069)
Philippines	0.157	(0.064)	-2.269	(0.068)	-0.561	(0.069)	-2.166	(0.069)	-0.638	(0.062)	-0.408	(0.063)	-0.030	(0.064)	-0.771	(0.063)
Poland	1.837	(0.082)	0.320	(0.076)	0.478	(0.073)	-0.216	(0.074)	-1.711	(0.077)	0.963	(0.071)	1.004	(0.069)	1.365	(0.073)
Portugal	0.291	(0.074)	0.035	(0.072)	0.299	(0.072)	0.043	(0.074)	0.351	(0.072)	0.143	(0.068)	1.379	(0.067)	0.302	(0.071)
Russia	1.162	(0.067)	-0.474	(0.064)	2.046	(0.063)	0.024	(0.061)	-1.057	(0.060)	1.794	(0.063)	1.940	(0.062)	1.173	(0.063)
Sweden	-0.822	(0.081)	-0.685	(0.076)	-0.980	(0.077)	-1.040	(0.072)	0.921	(0.071)	-1.592	(0.073)	-0.948	(0.070)	-1.293	(0.075)
Slovenia	0.460	(0.074)	-1.596	(0.076)	0.700	(0.072)	-0.574	(0.072)	-0.868	(0.072)	-0.445	(0.069)	0.088	(0.068)	0.039	(0.067)
Slovakia	0.340	(0.070)	-0.271	(0.067)	0.282	(0.066)	0.105	(0.075)	0.187	(0.064)	0.007	(0.064)	0.901	(0.064)	-0.065	(0.067)
Turkey	0.287	(0.068)	-0.877	(0.066)	-0.181	(0.067)	-0.465	(0.066)	-0.515	(0.063)	-0.874	(0.064)	-0.210	(0.065)	-0.076	(0.062)
South-Africa	0.130	(0.055)	-1.339	(0.054)	0.809	(0.053)	0.006	(0.052)	-0.212	(0.050)	-0.833	(0.051)	0.765	(0.050)	-0.416	(0.050)

Chapter 5

Including Covariates in Two-Stage Unobserved Consideration Set Discrete Choice Models

5.1 Introduction

Discrete choice modeling stems from the seminal work of Daniel McFadden (see, among many others, McFadden, 1978) and is widely practiced in empirical literature. Imagine the choice among various products in a supermarket where interest lies in the effect of pricing and advertisement. Standard models (Franses and Paap, 2001) assume that all individuals take all possible choice options into account. It is nonetheless likely that individuals choose from an individual-specific subset of the outcome space (Gensch, 1987). That is, individual i might only consider the cheapest products while individual j is interested in the low fat brands. Hence, the choice problem is seen as a two-step procedure, where first the consideration set is formed and then the final choice is made from the consideration set. These individual-specific considerations are typically unobserved by the researcher.

There is a vast literature on this two-step procedure originating with Howard and Sheth (1969). Wright and Barbour (1977) and Hauser and Wernerfelt (1990) describe economic theory and background. Sheridan et al. (1975) give an early application of a two-step procedure in job choice and Swait (1984) applies the theory to transportation demand. Since individual-specific considerations are often not observed, they are handled as latent variables in the estimation procedure (Manrai, 1995). Maximum likelihood estimation (ML, Ben-Akiva and Boccara, 1995) and Bayesian routines (van Nierop et al., 2010) are widely used. Compared to a benchmark multinomial logit model, Andrews and Srinivasan (1995) show significant improvements in terms of model fit of incorporating unobserved considerations to the choice process. Further, effect sizes of explanatory variables differ and are typically larger if the two-stage approach is taken. For an overview of the vast literature of two-step choice modeling we refer to Roberts and Lattin (1997).

The current chapter adds a note to the existing literature. Since inducing a consideration set in the choice model introduces two stages, an important question is which covariates impact which stage of the choice process. If considerations were observed, one could look to variable selection in the consideration stage. This is impossible with unobserved considerations and classification of covariates is therefore more difficult. The division of covariates over the consideration and final choice stage is however important for interpretation. For example, do price reductions of a supermarket brand change individual-specific consideration probabilities or do they directly affect choice? Andrews and Srinivasan (1995) include in their brand choice application all covariates in both choice stages and do not give a justification for this choice. Bronnenberg and Vanhonacker (1996) assume that prices are local (within the consideration set) instead of global. van Nierop et al. (2010) assume that "consideration is driven by in-store merchandising activity", and hence assume that "these variables do not directly drive [choice] utility".

The main research questions of this chapter are whether such theoretical assumptions are needed for the division of covariates over the stages and whether misplacement considerably affects in- and out-of-sample fit. That is, if it is not certain how to divide the covariates, can optimal classification be attained from model fit? Further, are obtained consideration probabilities informative if the division of covariates in the two-stage choice problem is unknown?

The outline of this chapter is as follows. First the model specification and parameter inference is described in Section 5.2. The description is only brief, since we use a basic model from the two-stage modeling literature (Swait, 1984). Section 5.3 presents a Monte Carlo study to investigate the impact of wrong placement of covariates in the two stages of the discrete choice model. We investigate in- and out-of-sample fit, consideration probabilities and parameter estimates. Finally, Section 5.4 concludes.

5.2 Model Specification

To model both unobserved consideration and multinomial choices we consider two stages of the decision process: a consideration and a decision stage. This section describes the model specification, where we assume that the (unobserved) considerations of choice items follow a binary logit model and a multinomial logit specification identifies the decision stage (Swait, 1984).

Consideration stage

Let C_{ij} denote the random variable describing the $0/1$ consideration decision for individual $i = 1, \ldots, N$ for choice option $j = 1, \ldots, J$. Note, these C_{ij} are unobserved by the econometrician and equal 1 if item j is in the consideration set of individual i and 0 otherwise. We describe this random variable by a binary logit specification given by

$$\Pr[C_{ij} = 1 | W_{ij}, \gamma_j, \delta_j] = \frac{\exp(\gamma_j + W_{ij}\delta_j)}{1 + \exp(\gamma_j + W_{ij}\delta_j)}, \tag{5.1}$$

where γ_j is an intercept and where W_{ij} is a choice-specific set of explanatory variables with corresponding parameter vector δ_j.

We assume that the random C_{ij} are independent and hence the joint distribution of $C_i = (C_{i1}, \ldots, C_{iJ})$ is given by

$$\Pr[C_i = c_i | W_i, \theta] = \prod_{j=1}^{J} \frac{\exp(c_{ij}(\gamma_j + W_{ij}\delta_j))}{1 + \exp(\gamma_j + W_{ij}\delta_j)}, \tag{5.2}$$

where θ summarizes the model parameters and c_i is a possible realization from the complete outcome space S. This outcome space contains all possible consideration sets and its size thus equals 2^J.

Choice stage

Let Y_i be a multinomial random variable describing the choice of individual i which can take the values $0, \ldots, J$. These $J + 1$ choices correspond to J choice items and a no choice option. This no choice option is added to allow for an empty consideration set from the first stage. The multinomial decision problem given the consideration set c_i is described by a multinomial logit model, that is,

$$\Pr[Y_i = j | c_i, X_{ij}, \theta] = \frac{c_{ij} \exp(\alpha_j + X_{ij}\beta_j)}{\sum_{l=0}^{J} c_{il} \exp(\alpha_l + X_{il}\beta_l)}, \tag{5.3}$$

where α_j is the choice-specific intercept of the model and where X_{ij} is a choice-specific set of explanatory variables with corresponding parameter vector β_j. The model is conditional on the first stage and hence if individual i does not consider item j, the probability of choosing j equals 0.

To estimate the parameters θ of the two-stage choice process we consider the maximum likelihood approach. This log-likelihood function is given by

$$\begin{aligned} \ell(\theta; y) &= \sum_{i=1}^{N} \log \Pr[Y_i = y_i] \\ &= \sum_{i=1}^{N} \log \sum_{c_i \in S} \Pr[Y_i = y_i | c_i, X_i, \theta] \Pr[C_i = c_i | W_i, \theta], \end{aligned} \tag{5.4}$$

where $\Pr[C_i = c_i|W_i, \theta]$ and $\Pr[Y_i = y_i|c, X_i, \theta]$ are given in (5.2) and (5.3), respectively. Note that the consideration sets are unobserved and the log-likelihood function therefore consists of a weighted sum over all possible consideration sets. That is, the model actually is a concomitant variable mixture type model (see Wedel, 2002) with restrictions on the parameters. Consideration probabilities that come out of the estimation procedure give insights on the unobserved considerations.

As the two stages of the decision process interact, it is probably not easy to determine which covariate affects which stage. In the supermarket example, the price of the product might influence consideration, the final product choice, or both. In the next section we investigate via Monte Carlo experiments whether we can find a justification for the division of covariates over the two stages via statistical model comparison. If we cannot find such justification, the division of covariates over W_{ij} and X_{ij} should be based on solid theoretical grounds. Furthermore, we investigate the impact of the division of covariates on the inference of the consideration probabilities.

5.3 Simulation Study

In this section we conduct several simulation studies to investigate the effect of covariate misplacement in the two-stage choice model described in Section 5.2. We answer the research questions from Section 5.1 in particular by looking at in- and out-of-sample fit, fitted consideration probabilities and parameter estimates.

We consider the model specification as in (5.2) and (5.3) where the number of choice alternatives J is 3. Hence, the individuals have 4 choice options since a no choice option is added to the choice stage. We consider an empirically likely sample size N of 1000 and do 5000 replications. As explanatory variables we take two standard normally distributed random variables, z_1 and z_2. These explanatory variables are used in three data generating processes [DGPs] as depicted in Table 5.1. The parameters of our DGPs are chosen such that large, moderate and small consideration and choice probabilities occur, see Table 5.4 for the values of our DGP-parameters.

Table 5.1: DGPs and model specifications including abbreviations of the models used in the simulation study.

Data generating processes	
$m(z_1; z_2)$	z_1 impacts consideration, z_2 impacts choice
$m(z_1, z_2; --)$	z_1 and z_2 impact consideration
$m(--; z_1, z_2)$	z_1 and z_2 impact choice
Model specifications	
$m(z_1; z_2)$	z_1 impacts consideration, z_2 impacts choice
$m(z_2; z_1)$	z_2 impacts consideration, z_1 impacts choice
$m(z_1, z_2; --)$	z_1 and z_2 impact consideration
$m(--; z_1, z_2)$	z_1 and z_2 impact choice
$m(z_1, z_2)$	benchmark multinomial logit model, z_1 and z_2 impact choice

Interest lies in the impact of misplacement of the explanatory variables. That is, if the researcher in the supermarket example erroneously puts the price of the product in the consideration stage instead of the choice stage, what would be the impact on model fit? Furthermore, we address whether interpretation on the latent consideration probabilities will be correct. We investigate these questions by introducing five model specifications shown in Table 5.1 for the three different DGPs. The simulation study is highly stylized and extensions would only increase difficulty in the classification of covariates.

To compare the estimation results of the five models, we investigate in- and out-of-sample hit rate and log-likelihood values of all model specifications for each DGP. We define the hit rate as the percentage of observations where the model specification gives largest probability to the actually chosen alternative. We investigate the hit rate of both consideration and choice probabilities. Note that in empirical studies the consideration probability is unobserved and a hit rate cannot be found.

We compare the (out-of-sample) hit rates of the different model specifications by looking at the distribution over the replications. Further, we look at the percentage of replications where the correct model specification outperforms the other specifications. The parametric Vuong (1989) test can be used to compare log-likelihood values of non-nested model specifications. If the hit rates and the Vuong test point to the model specification in line with the DGP, classification of covariates can be based on these measures. Otherwise, solid theoretical grounds are needed to identify which covariate enters which stage of the decision process in the unobserved consideration set model.

5.3.1 Simulation Results

This section presents the results of the simulation studies. First, we evaluate the fit of the unobserved consideration sets in the misspecified model specification. Second, we present parameter estimates. Finally, we investigate in- and out-of-sample hit rates and log-likelihood values over the replications. Hence, this is the core of our research. If hit rates and log-likelihood values are clearly optimal for the correct specification, the division of the covariates over the consideration and choice stage can be based on these statistical measures.

Consideration Sets

Table 5.2 displays the average size of the 2^J estimated consideration sets for all specifications. Note that the comparison to the actual unobserved consideration sets cannot be done in empirical studies. The table shows accurate sizes of the consideration sets for the correctly specified model. Nonetheless, large deviations from the simulated values for the misspecified model specifications occur. For example, for model specification $m(z_1, z_2; --)$ in the study with DGP $m(z_1; z_2)$, 30$\%$ of the observations are in the consideration set $(0, 1, 0)$, although the DGP corresponds to a probability of only 6$\%$ for this set. Apparently, the consideration stage probabilities depend highly on which covariate enters which stage of the choice problem. Table 5.3 shows the average hit rate of the considerations. Clearly, these hit rates are small if the division of covariates over the two stages is incorrect. Where in DGP $m(z_1; z_2)$ the in- and out-of-sample hit rate is near 50$\%$ for correct classification, the other hit rates are significantly worse. Wrong classification in the two-stage decision process thus yields highly different estimates for the consideration probabilities. Hence, conclusions on these consideration predictions can only be valid if the division of covariates is very certain to be correct.

Table 5.2: Average size of the estimated consideration sets over the replications for each model specification in all DGPs (1000 observations, 5000 replications).

DGP Consideration set	$m(z_1; z_2)$ DGP		Model Specification							
			$m(z_1; z_2)$		$m(z_2; z_1)$		$m(z_1, z_2; --)$		$m(--; z_1, z_2)$	
	mean	std[a]	mean	std	mean	std	mean	std	mean	std
000[b]	0.116	0.010	0.117	0.036	0.043	0.027	0.320	0.039	0.009	0.011
001	0.008	0.003	0.009	0.006	0.170	0.088	0.039	0.014	0.051	0.050
010	0.061	0.008	0.065	0.032	0.131	0.039	0.308	0.041	0.016	0.018
011	0.021	0.005	0.026	0.020	0.454	0.096	0.166	0.024	0.175	0.097
100	0.286	0.015	0.280	0.048	0.032	0.020	0.065	0.016	0.024	0.027
101	0.046	0.007	0.045	0.017	0.016	0.011	0.004	0.002	0.155	0.132
110	0.268	0.014	0.267	0.047	0.104	0.023	0.076	0.019	0.044	0.042
111	0.194	0.012	0.192	0.031	0.050	0.014	0.021	0.007	0.527	0.188
DGP: Consideration set	$m(z_1, z_2; --)$ DGP		Model Specification							
			$m(z_1; z_2)$		$m(z_2; z_1)$		$m(z_1, z_2; --)$		$m(--; z_1, z_2)$	
	mean	std	mean	std	mean	std	mean	std	mean	std
000	0.152	0.012	0.146	0.029	0.054	0.028	0.153	0.025	0.049	0.022
001	0.040	0.006	0.038	0.015	0.115	0.042	0.041	0.011	0.142	0.059
010	0.071	0.008	0.142	0.040	0.080	0.038	0.073	0.020	0.076	0.044
011	0.071	0.008	0.195	0.056	0.131	0.039	0.072	0.018	0.196	0.085
100	0.285	0.014	0.151	0.038	0.166	0.084	0.284	0.046	0.060	0.033
101	0.023	0.005	0.037	0.021	0.065	0.033	0.023	0.010	0.182	0.105
110	0.267	0.014	0.130	0.031	0.295	0.086	0.266	0.043	0.080	0.035
111	0.091	0.009	0.160	0.044	0.094	0.033	0.089	0.018	0.214	0.055
DGP: Consideration set	$m(--; z_1, z_2)$ DGP		Model Specification							
			$m(z_1; z_2)$		$m(z_2; z_1)$		$m(z_1, z_2; --)$		$m(--; z_1, z_2)$	
	mean	std	mean	std	mean	std	mean	std	mean	std
000	0.098	0.009	0.079	0.037	0.168	0.083	0.095	0.041	0.095	0.037
001	0.036	0.006	0.026	0.020	0.176	0.116	0.192	0.112	0.037	0.016
010	0.098	0.009	0.117	0.072	0.231	0.130	0.090	0.050	0.098	0.043
011	0.036	0.006	0.050	0.038	0.195	0.105	0.368	0.118	0.037	0.016
100	0.267	0.014	0.147	0.080	0.055	0.036	0.044	0.022	0.262	0.053
101	0.098	0.009	0.063	0.061	0.038	0.035	0.040	0.038	0.103	0.031
110	0.267	0.014	0.340	0.127	0.089	0.042	0.065	0.027	0.266	0.054
111	0.098	0.009	0.178	0.106	0.048	0.025	0.106	0.037	0.102	0.021

[a] The mean represents the mean over 5000 replications of the size of the consideration sets within each replication. The standard deviation over the replications is represented by std.

[b] A consideration set like 101 represents the set where choice option 1 and 3 are considered and choice option 2 is not.

Table 5.3: Average hit rate of consideration over the replications for each model specification in all DGPs (1000 observations, 5000 replications).[a]

DGP	$m(z_1; z_2)$			
Model Specification	In-sample hit rate		Out-of-sample hit rate	
	mean	std[b]	mean	std
$m(z_1; z_2)$	0.474	0.019	0.470	0.019
$m(z_2; z_1)$	0.062	0.011	0.062	0.011
$m(z_1, z_2; --)$	0.164	0.017	0.163	0.017
$m(--; z_1, z_2)$	0.184	0.040	0.184	0.041
DGP	$m(z_1, z_2; --)$			
Model Specification	In-sample hit rate		Out-of-sample hit rate	
	mean	std	mean	std
$m(z_1; z_2)$	0.219	0.047	0.216	0.046
$m(z_2; z_1)$	0.292	0.025	0.291	0.025
$m(z_1, z_2; --)$	0.528	0.021	0.523	0.021
$m(--; z_1, z_2)$	0.057	0.029	0.056	0.029
DGP	$m(--; z_1, z_2)$			
Model Specification	In-sample hit rate		Out-of-sample hit rate	
	mean	std	mean	std
$m(z_1; z_2)$	0.237	0.048	0.235	0.048
$m(z_2; z_1)$	0.105	0.031	0.104	0.031
$m(z_1, z_2; --)$	0.076	0.015	0.075	0.014
$m(--; z_1, z_2)$	0.269	0.019	0.266	0.019

[a] The hit rate is defined as the fraction of observations where the model specification gives largest probability to the actually considered consideration set.

[b] The mean represents the mean over 5000 replications of the hit rate of consideration. The standard deviation over the replications is represented by std.

Parameter Estimates

Table 5.4 displays the average parameter estimates of the correctly specified model specifications. Parameter estimates of the misspecified models are highly different, since the consideration probabilities are highly different from the data generating process. The table shows that parameters are accurately estimated and reasonable root mean squared errors are found.[1] Only the estimates of γ_j in $m(--;z_1,z_2)$ are somewhat skewed. Tables 5.2 and 5.3 nonetheless showed that consideration probabilities are estimated rather accurately.

Table 5.4: Average parameter estimates and root mean squared error over the replications for the correct model specification in all DGPs (1000 observations, 5000 replications).[a]

	$m(z_1;z_2)$				$m(z_1,z_2,--)$				$m(--;z_1,z_2)$		
	DGP	mean	rmse		DGP	mean	rmse		DGP	mean	rmse
α_1	0	–	–	α_1	0	–	–	α_1	0	–	–
β_1	0	–	–	α_2	0.5	0.523	0.180	β_1	0	–	–
α_2	-1.75	-1.764	0.236	α_3	1	1.072	0.696		0	–	–
β_2	-2.25	-2.337	0.317	α_4	2.25	2.365	0.965	α_2	-1	-1.002	0.281
α_3	1	1.056	0.475	γ_1	1.25	1.336	0.562	β_2	0.75	0.785	0.207
β_3	0	0.034	0.259	δ_1	0.75	0.795	0.312		-2.25	-2.358	0.392
α_4	2.5	2.610	0.669		-2.25	-2.378	0.532	α_3	2	2.194	0.869
β_4	2.25	2.396	0.736	γ_2	0	0.030	0.398	β_3	1.5	1.627	0.476
γ_1	1.5	1.717	1.160	δ_2	1.5	1.565	0.340		0	0.031	0.292
δ_1	0.75	0.893	0.660		0	-0.008	0.156	α_4	3.5	3.794	1.280
γ_2	0.25	0.314	0.375	γ_3	-2.75	-2.790	0.251	β_4	2.25	2.468	0.704
δ_2	1.5	1.565	0.306	δ_3	2.25	2.299	0.334		2.25	2.550	1.041
γ_3	-1.75	-1.763	0.167		2.25	2.294	0.299	γ_1	1	1.443	2.880
δ_3	2.25	2.313	0.251					γ_2	0	0.084	1.230
								γ_3	-1	-0.956	0.189

[a] For reasons of space, the parameter estimates of the other model specifications are available upon request.

[1] Some replications in the simulation study converge to parameter estimates where a consideration probability approaches 1. Since parameter estimates are then not clearly identified, we discard these replications from the analysis. Note that this extreme convergence happens more often for the misspecified than for the correctly specified models.

Hit rate and Log-likelihood

Given the conclusions from the previous paragraphs, it is important to classify the division of covariates carefully. Both unobserved consideration probabilities and parameter estimates are highly different if the model specification is not in accordance with the DGP. The main question of this research is whether the classification of covariates can be done based on statistical measures. Since the size of the estimated consideration sets cannot be compared to unobserved consideration sets in empirical research, this (clear identification) can unfortunately not be used. Table 5.5 displays the average hit rate of the choices for each model specification. The hit rates are very close to each other for all data generating processes. That is, from these hit rates only it is not possible to identify which model specification is in correspondence with the DGP. The table further shows the fraction of replications where the correct model specification has the highest in- and out-of-sample hit rate. Since in DGP $m(z_1; z_2)$ this is only the case in 52 and 46$\%$, respectively, it is not save to base the division of covariates on this statistical measure. That is, the wrong specification would have been chosen in roughly 50$\%$ of the replications. Hence, Table 5.2 showed that consideration probability estimates are then far from the truth and cannot be interpreted savely.

Thus, decisions on classification of covariates cannot be based on in- and out-of-sample hit rates. We therefore introduce parametric testing. First, practitioners might look at Wald tests for the inclusion of covariates. Unfortunately, unreported test results show that this test cannot discriminate between the model specifications in Table 5.1. This is not surprising, since both explanatory variables in the end influence final choice. Although this influence might by direct instead of indirect, the optimization will still try to capture relations between the covariate and discrete choice. Hence, significance of the parameters are still found in the misspecified models. Next, we opt for the Vuong test for comparison of non-nested models. Table 5.6 shows the fraction of Z-scores over the replications larger than 1.645, thus reporting the fraction of rejection of the null hypothesis that the model specifications are equally likely at a 5$\%$ level. For example, $m(z_1; z_2)$ significantly outperforms $m(z_2; z_1)$ in only 39,5$\%$ of the replications, while $m(z_1; z_2)$ is in accordance with the DGP. That is, we would not be able to identify the correct specification in 60$\%$ of the cases. We cannot decide on classification of covariates in the two-stage problem based on this Vuong tests. Only DGP $m(z_1, z_2, --)$ is significantly identified since the Vuong test punishes the other specifications for the larger number of parameters.

Table 5.5: Average hit rate of choice over the replications for each model specification in all DGPs (1000 observations, 5000 replications).[a]

DGP	$m(z_1; z_2)$			
Model Specification	In-sample hit rate		Out-of-sample hit rate	
	mean	std[b]	mean	std
$m(z_1; z_2)$	0.569	0.016	0.564	0.016
$m(z_2; z_1)$	0.567	0.016	0.562	0.016
$m(z_1, z_2; --)$	0.561	0.016	0.557	0.016
$m(--; z_1, z_2)$	0.562	0.016	0.558	0.016
$m(z_1, z_2)$	0.561	0.016	0.558	0.016
fraction[c]	0.521		0.462	
DGP	$m(z_1, z_2; --)$			
Model Specification	In-sample hit rate		Out-of-sample hit rate	
	mean	std	mean	std
$m(z_1; z_2)$	0.479	0.016	0.469	0.017
$m(z_2; z_1)$	0.484	0.016	0.476	0.016
$m(z_1, z_2; --)$	0.495	0.016	0.488	0.016
$m(--; z_1, z_2)$	0.487	0.017	0.480	0.017
$m(z_1, z_2)$	0.483	0.016	0.479	0.016
fraction	0.665		0.617	
DGP	$m(--; z_1, z_2)$			
Model Specification	In-sample hit rate		Out-of-sample hit rate	
	mean	std	mean	std
$m(z_1; z_2)$	0.433	0.016	0.424	0.016
$m(z_2; z_1)$	0.432	0.016	0.424	0.016
$m(z_1, z_2; --)$	0.429	0.016	0.423	0.016
$m(--; z_1, z_2)$	0.435	0.015	0.429	0.016
$m(z_1, z_2)$	0.430	0.016	0.425	0.016
fraction	0.347		0.369	

[a] The hit rate is defined as the fraction of observations where the model specification gives largest probability to the actual choice.
[b] The mean represents the mean over 5000 replications of the hit rate of choice. The standard deviation over the replications is represented by std.
[c] Fraction denotes the fraction of replications where the correct model specification has the highest hit rate.

Table 5.6: Fraction of Z-scores of the Vuong test larger than 1.645 over the replications for each model specification in all DGPs (1000 observations, 5000 replications).[a]

DGP Model Specification	$m(z_1; z_2)$ fraction[b]	$m(z_1, z_2, --)$ fraction	$m(--; z_1, z_2)$ fraction
$m(z_1; z_2)$	–	0.999	0.566
$m(z_2; z_1)$	0.395	0.991	0.409
$m(z_1, z_2; --)$	0.790	–	0.009
$m(--; z_1, z_2)$	0.972	0.958	–
$m(z_1, z_2)$	0.670	0.967	0.304
Fraction of log-likelihoods[c]			
In-sample	0.927	0.949	0.957
Out-of-sample	0.878	0.932	0.902

[a] The Z-score is the test statistic from the parametric Vuong test comparing likelihood values of non-nested models.
[b] Fraction denotes the fraction of replications where the Z-score is larger than 1.645, thus where the Vuong test significantly rejects the null hypothesis at a 5 percent level in favour of the correct model specification.
[c] Fraction of log-likelihoods denotes the fraction of replications where the correct model specification has the highest log-likelihood value.

Both the hit rate and Vuong test cannot give a certain classification of covariates in the two-stage choice model. We finally look at log-likelihood values. Table 5.6 shows the in- and out-of-sample fraction of replications where the correct model specification has the largest log-likelihood value. These measures give clearer judgment on which covariate has to enter which stage of the decision process. That is, the log-likelihood value of the specification in accordance with the DGP is largest in more than 90% of the replications.

In sum, the division of covariates over the two stages in the choice process is important for interpretation of the estimates of the unobserved consideration probabilities. In- and out-of-sample hit rates give no definite answer to the placement question, since hit rates of the specification in accordance of the DGP do not clearly outperform the other specification hit rates. Further, the Vuong test on the log-likelihood values will not yield the correct specification with enough certainty, either. The choice on which covariate enters which stage of the decision process can best be based on log-likelihood values: the largest log-likelihood value indeed belongs to the correct specification in about 90% of the replications. Since some uncertainty is still involved, theoretical grounds for classification are recommended. Note that the classification of covariates does not matter much for in- and out-of-sample fit of the final choices. Nonetheless, it is highly important for parameter interpretation and interpretation of the unobserved consideration probabilities.

5.4 Conclusion

In this chapter we provide a critical note to the two-stage unobserved consideration set literature. In this literature, it is assumed that the choice process can be divided into two stage. First, an individual-specific and unobserved subset of choice options is created in the consideration stage. Second, the choice is made from this subset of options. There is vast theoretical evidence that individual choices are made in this way instead of in the standard one-stage choice process.

We note that implications arise when the inclusion of covariates is concerned. That is, the division of covariates over the two stages is highly important for interpretation of unobserved consideration probabilities and parameter estimates. Nonetheless, effects on in- and out-of-sample fit of the choice variable are only small and negligible and do not help to identify covariate classification. If placement of the covariates cannot be based on statistical measures, solid theoretical grounds are needed to justify interpretation of the model parameters.

We find that both in- and out-of-sample hit rate and the parametric Vuong test do not give clear justification for the classification of covariates. If one wants to decide upon inclusion of covariates in the two-stage decision problem based on statistical measures, it is best to choose the specification with the largest log-likelihood value. Then, the correct model specification is chosen in more than 90% of the replications of the simulation study. Note that in 10%, the wrong specification would have been chosen, and interpretation of consideration probabilities would have been far from the truth. Thus, theoretical grounds are still highly recommended.

More information in the modeling process would give more justification for covariate inclusion. That is, if the number of observations N is larger, identification of the correct model specification would be easier. We choose N equal to 1000 since this is a sample size likely to encounter in practice. Further, if individuals are repeatedly measured over time and structure is laid on the dynamic consideration probabilities, clearer conclusions on the classification of covariates may be found. Nonetheless, the explanatory variables in the current study are uncorrelated and correlations would yield extra difficulties to discriminate between the model specifications. Note, it is always uncertain whether the model specification is in accordance with the DGP. Unobserved consideration probabilities should therefore be treated with care.

Chapter 6

Modeling the Impact of Forecast-based Regime Switches on US Inflation

6.1 Introduction

Lucas (1976) showed that macro-econometric models with constant parameters cannot be used for evaluating policy changes, since policy changes usually result in behavioral changes of economic agents. Hence, these behavioral changes result in inconstant model parameters. It is well known that agents also react to macro-economic forecasts. This suggests that unexpected economic forecasts may also lead to changes in the model parameters over time.

Several theoretical and empirical studies indicate this effect of forecasts. Theoretically, Fellner (1976) explains that publics expectations are prone to self-justifying skepticism about policy makers and policy makers react to that. Empirically, Givoly and Lakonishok (1979) find that serious upward revisions in financial earnings forecasts lead to significant effects on stock prices. Steiner et al. (2009) show that macro-economic announcements cause an immediate reaction of returns in asset prices. Moreover, they find that reactions to positive news are faster than reaction to negative news. Sinclair et al. (2012) shows that forecast errors have an impact on the target interest rate set by the Federal Reserve Bank [FED].

Although literature suggests that the impact of forecasts occurs in various fields, this chapter mainly focuses on US inflation time series data. The choice for this series is coherent, as (i) policy makers react to forecasts due to the policy of inflation targeting starting in the FED Volcker-regime in 1975 (Clarida et al., 2000) and; (ii) companies and consumers use inflation forecasts to decide upon future savings and expenditures. Carroll (2003) states that people update their expectations to public forecasts rather than to past inflation rates.

Furthermore, economic theory also provides support for the impact of inflation forecasts on the inflation rate. It is mainly mentioned as the expectations trap (Christiano and Gust, 2000) or self-fulfilling expectations, where publics expectations of high inflation increase the actual inflation rate. Albanesi et al. (2003) state: "expectations of high or low inflation lead the public to take defensive actions, which then make accommodating those expectations the optimal monetary policy". Both the expectations trap before 1979 (Leduc et al., 2007) and inflation targeting since the 1980s suggest that inflation forecasts play a key role.

To describe the effect of forecasts on future inflation, we propose in this chapter a nonlinear time series model which accounts for dynamic effects of (inflation) forecasts. That is, the model allows for structural breaks in the parameters based on the relative size of a forecast of the underlying time series. To describe these structural changes we employ the smooth transition autoregressive [STAR] model (Chan and Tong, 1986; Teräsvirta and Anderson, 1992). Although there are many applications of regime switching models, none of these consider the impact of forecasts on regime changes. In many time series applications the transition is based on a lagged value of the dependent variable, see Teräsvirta (1994), among many others. Now, the classification into regimes depends on a forecast of the dependent variable. Note that evaluation of the forecast is impossible due to the Lucas (1976) critique. This forecast may either be exogenous in the sense that it is formed outside the model or endogenous when the forecast is generated inside the model specification. In the latter case, the proposed model resembles the contemporaneous STAR [C-STAR] model of Dueker et al. (2007) and hence provides a motivation for this specification.

The remainder of this chapter is organized as follows. Section 6.2 introduces our model specification to describe the impact of forecasts. Parameter estimation, statistical inference and a nonlinearity test are discussed in Section 6.3. We perform several simulation studies to justify the validity of the test. Section 6.4 illustrates our modeling approach using the gross domestic product [GDP] deflator based inflation rate of the United States [US]. Finally, Section 6.5 concludes.

6.2 Model Specification

We put forward a nonlinear time series model which accounts for structural changes due to a forecast of the underlying time series. As we expect reactions to relatively low and relatively high forecasts, we include three regimes. We further expect that the size of the structural change depends on the size of the forecast. Therefore, we use smooth transition modes, see van Dijk et al. (2002) for a survey.

Formally, let y_t be the variable of interest at time $t = 1, \ldots, T$. Let $p_{t|t-1}$ denote the forecast of y_t based upon all information up to and including time $t-1$. The three-regime smooth transition time series then is given by

$$y_t = \phi_1 x_t + (\phi_0 - \phi_1) x_t G_0(p_{t|t-1}; \gamma_0, \kappa_0) + (\phi_2 - \phi_1) x_t G_2(p_{t|t-1}; \gamma_2, \kappa_2) + \sigma_t \varepsilon_t \tag{6.1}$$

with $\varepsilon_t \sim IID(0, 1)$, where x_t is a k-dimensional vector containing a vector of ones, explanatory variables and lagged values of y_t and where ϕ_i, $i \in \{0, 1, 2\}$, are $(k \times 1)$-parameter vectors. The parameter σ_t describes the potentially time-varying standard deviation of the disturbances which we will discuss later.[1]

The functions $G_0(\cdot)$ and $G_2(\cdot)$ takes values between 0 and 1 depending on the level of the forecasts $p_{t|t-1}$ and describe the probability to be below or above some threshold value. We opt for the logistic function

$$G_i(p_{t|t-1}; \gamma_i, \kappa_i) = \frac{1}{1 + \exp(-\gamma_i(p_{t|t-1} - \kappa_i))}, \tag{6.2}$$

resulting in the logistic STAR [L-STAR] model (Teräsvirta, 1994). The parameter γ_i determines the smoothness of the transition function and κ_i denotes he point of inflection of the logistic curve. It is easy to see that under the restriction $\kappa_0 < \kappa_2$, $\gamma_0 < 0$ and $\gamma_2 > 0$, $G_0(\cdot)$ approaches 1 for small forecasts. Hence, the relevant parameter vector is ϕ_0. For large forecasts, $G_2(\cdot)$ approaches 1 resulting in ϕ_2 as the relevant parameter vector. These restrictions are however not necessary for identification but other restrictions may lead to different interpretation of the regime parameters.

6.2.1 Specification of the influential forecast

In many time series applications of STAR models $p_{t|t-1}$ is replaced by y_{t-1}. To serve the purpose of our model we take a different approach. The classification into regimes depends on the forecast $p_{t|t-1}$ of the dependent variable y_t. Different specifications can be used. The impact of the forecast $p_{t|t-1}$ should be important enough to result in reactions of decision makers in the economy.

The simplest case is if the forecast stems from an expert opinion or from another econometric model. In this case we obtain a regular L-STAR model. One can also use the model in (6.1) to provide the forecast. Then, we assume that the forecaster is familiar with the impact of his or her forecast and acts to that by incorporating regime changes. As the forecaster does not yet have knowledge on future regime switches, (s)he can do no better than imposing the state of time $t-1$ in his or her naive

[1] For now, the variance is independent of changes in the forecast $p_{t|t-1}$.

forecast. The relevant forecast for period t given the information at time $t-1$ is therefore given by

$$p_{t|t-1} = \phi_1 x_{t|t-1} + (\phi_0 - \phi_1) x_{t|t-1} G_0(p_{t-1|t-2}; \gamma_0, \kappa_0) + (\phi_2 - \phi_1) x_{t|t-1} G_2(p_{t-1|t-2}; \gamma_2, \kappa_2), \quad (6.3)$$

where we use the previous realization of the transition functions $G_i(\cdot)$ and where $x_{t|t-1}$ is the set of explanatory variables for time t given all information up to time $t-1$.

The model specification (6.1)-(6.3) adopts and extends the ideas of Dueker et al. (2007). They propose a STAR model with contemporaneous classification. Since $p_{t|t-1}$ contains model parameters it belongs to the class of contemporaneous models. Our current representation of this C-STAR model provides a justification and interpretation of using a contemporaneous, not predetermined classification into regimes. Further, we extend the model of Dueker et al. (2007) from two to three regimes.

6.2.2 Specification of the time-varying threshold

The threshold parameter κ_i in (6.2) is assumed to be fixed in original STAR specifications. As macro-economic time series have been fluctuating in the past decades, it is nonetheless likely that reactions to the forecast are relative to time. For instance, a high forecast in the 1990s with a low inflation level would not have been striking during the oil crises in the late 1970s. We therefore allow the threshold to be time-varying, relative to the local level of inflation. That is, agents compare the forecast to the level of inflation series in the near past.

We consider two specifications for the time-varying κ_{it}. First, let $\kappa_{it} = \kappa_i + \bar{y}_t^{(d)}$, where $\bar{y}_t^{(d)}$ is the average of the dependent variable over the previous d periods. The larger $\bar{y}_t^{(d)}$, the larger $p_{t|t-1}$ has to be for agents to react. It also implies that regime 0 is more likely to occur. For the second specification of κ_{it}, imagine a large forecast in a highly volatile period. As large changes are expected, it is likely that reactions to this forecast are less extreme than reaction to the same forecast in periods with low volatility. We therefore impose that $\kappa_{it} = \kappa_i \sigma_t + \bar{y}_t^{(d)}$. Hence, we now also account for the local level of the variance in the inflation innovations.

In sum, the specification in (6.1)-(6.2) where $G_0(\cdot)$ and $G_2(\cdot)$ depend on the level of the forecast $p_{t|t-1}$ provides the framework for investigating the impact of forecasts on decisions of agents. We allow for exogenous and endogenous forecasts as described in (6.3) and for time-varying threshold parameters. The model allows us to investigate the impact of forecasts on macro-economic variables of interest.

6.3 Statistical Inference

We discuss inference of our smooth transition model specification from Section 6.2. Section 6.3.1 considers parameter estimation, while Section 6.3.2 concerns testing for our specific form of nonlinearity

6.3.1 Estimation procedure

To estimate the parameters in (6.1)-(6.2) we use nonlinear least squares [NLS], see, for example, Davidson and MacKinnon (2004, Chapter 6). It is however not possible to apply regular NLS procedures that are used for STAR models. First of all, the argument of the logistic function in (6.2) may depend on parameters as in (6.3). Further, we want to allow for a time-varying variance σ_t^2. We use weighted NLS [WNLS] instead.

We allow for a time-varying variance because of the so-called great moderation. Many macroeconomic time series display a drop in volatility in the 1980s, see Kahn et al. (2002) and Summers (2005). We allow for these changes by adding a break to the variance σ_t^2.

To facilitate notation, we define

$$f(x_t;\theta) = \phi_1 x_t + (\phi_0 - \phi_1)x_t G_0(p_{t|t-1};\gamma_0,\kappa_{0t}) + (\phi_2 - \phi_1)x_t G_2(p_{t|t-1};\gamma_2,\kappa_{2t}) \tag{6.4}$$

where $\theta = (\phi_0, \phi_1, \phi_2, \gamma_0, \gamma_2, \kappa_0, \kappa_2)$ and hence (6.1) can be written as

$$y_t = f(x_t;\theta) + \sigma_t \varepsilon_t. \tag{6.5}$$

To capture the great moderation we follow Sensier and van Dijk (2004) and define

$$\sigma_t^2 = \sigma_1^2 + (\sigma_2^2 - \sigma_1^2)G_\sigma(t;\gamma_\sigma,\kappa_\sigma) + \eta_t. \tag{6.6}$$

In contrast to Sensier and van Dijk (2004) we allow for the possibility of a smoother transition using

$$G_\sigma(t;\gamma_\sigma,\kappa_\sigma) = \frac{1}{1+\exp(-\gamma_\sigma(t-\kappa_\sigma))} \tag{6.7}$$

which again is the logistic function. Hence for $\gamma_\sigma > 0$ the variance is σ_1^2 for the first part of the sample and σ_2^2 for the second part. The transitions is halfway at $t = \kappa_\sigma$ and γ_σ reflects the smoothness of the transition.

The WNLS procedure to estimate the model parameters θ can be summarized by the following five steps:

1. minimize $\sum_{t=1}^{T}(y_t - f(x_t;\theta))^2$ with respect to θ resulting in $\hat{\theta}_0$
2. compute the residuals $\hat{\varepsilon}_t = y_t - f(x_t;\hat{\theta}_0)$
3. use NLS on (6.6) replacing σ_t^2 by $\hat{\varepsilon}_t^2$
4. compute the fitted values of σ_t^2 using (6.6) resulting in $\hat{\sigma}_t^2$
5. minimize $\sum_{t=1}^{T}(\frac{1}{\hat{\sigma}_t}(y_t - f(x_t;\theta)))^2$ with respect to θ resulting in $\hat{\theta}$

The estimator is asymptotically normally distributed. The covariance matrix of the estimator can be computed using

$$\hat{\sigma}_\varepsilon^2 \left(\sum_{t=1}^{T} \frac{1}{\hat{\sigma}_t^2} \left(\frac{\partial f(x_t;\theta)}{\partial \theta}|_{\theta=\hat{\theta}} \right) \left(\frac{\partial f(x_t;\theta)}{\partial \theta}|_{\theta=\hat{\theta}} \right)' \right)^{-1}. \tag{6.8}$$

Diagnostic tests on the residuals (such as heteroskedasticity and serial correlation tests) can be done in a similar manner as for linear time series models. Since there are unidentified nuisance parameters under the null hypothesis of linearity we cannot use standard tests to compare our model to a linear specification. In the next section we propose a nonlinearity test based on Luukkonen et al. (1988) to test for our specific type of nonlinearity.

6.3.2 Nonlinearity test

The first step in the modeling process is to test for the presence of our proposed type of nonlinearity. As comparing our model specification (6.1) with a linear model specification leads to the problem of unidentified parameters under the null hypothesis, standard tests do not apply. Instead, we use a test by Luukkonen et al. (1988) which is based on the first-order Taylor expansion around $\gamma_i = 0$ of the logistic function $G_i(\cdot)$ in (6.2).

A first-order Taylor expansion of the restricted two-regime model version of (6.1) results in

$$y_t = \phi_1 x_t + \tilde{\beta}_0 x_t + \tilde{\beta}_1 x_t p_{t|t-1} + \sigma_t \varepsilon_t, \tag{6.9}$$

where

$$\begin{aligned} \tilde{\beta}_0 &= (0.5 - 0.25\gamma_0\kappa_0)(\phi_0 - \phi_1) \\ \tilde{\beta}_1 &= 0.25\gamma_0(\phi_0 - \phi_1). \end{aligned} \tag{6.10}$$

It is easy to see that if $\gamma_0 = 0$ or $\phi_0 = \phi_1$ the additional regime is not present in the specification. Hence, the nonlinearity test boils down to testing $\tilde{\beta}_1 = 0$ using a standard Wald or t-test with a standard distribution. For testing for an additional third regime we use the approach of van Dijk and Franses (1999).

Simulation study

If regime switches in the model in (6.1) are based on an exogenous forecast $p_{t|t-1}$ it fits in the framework of Luukkonen et al. (1988). However ϕ_i also emerges in $p_{t|t-1}$ when the endogenous forecast in (6.3) is used and it is not straightforward to implement the test. We replace $p_{t|t-1}$ by its fitted value from the model in (6.1), $\hat{p}_{t|t-1}$, such that the standard linear test on $\tilde{\beta}$ in (6.9) can still be used. To justify whether this strategy leads to proper inference, we perform several simulation studies.

Under the null hypothesis we take a simple linear autoregressive model of order 1, that is

$$y_t = \rho_0 + \rho_1 y_{t-1} + \nu_t \qquad \text{for} \qquad t = 1, \dots, T, \tag{6.11}$$

where ρ_0 and ρ_1 are parameters and $\nu_t \sim NID(0, \sigma_\nu^2)$. To investigate the impact of the autoregressive parameters on the test we consider ρ_1 equal to 0.2, 0.75 and 0.95. Moreover, we choose ρ_0 to be 0.8, 0.25 and 0.05, respectively, so that the unconditional mean of the time series equals 1. We compare the empirical size of the test for $\tilde{\beta} = 0$ in the test regression (6.9) with the nominal size.

Table 6.1 displays the empirical size of the test based on 10000 replications. Since we generate the data without any nonlinearity, empirical sizes give the probability to wrongly reject the null hypothesis. For autoregressive parameters not close to unit root rather small size distortions occur, even for 250 observations. For $\rho_1 = 0.95$, the size distortion is bigger but not severe. For example, for $T = 250$ the empirical size belonging to the significance level of 10% is about 5%, while for $\rho_1 = 0.75$ and $\rho_1 = 0.2$ the empirical size is about 8 and 9%, respectively.

All values in Table 6.1 are smaller than the corresponding theoretical size. This implies that the test is a bit too conservative. However, these small size distortions cause no severe problems in practice. Moreover, size distortions get smaller for larger T and hence the test seems to be asymptotically valid.

Unreported results show that similar results are found for the regular STAR model and our model with the exogenous forecast. This indicates that the nonlinearity test introduced by Luukkonen et al. (1988) is appropriate to use for the model specification in (6.1)-(6.3).

Table 6.1: Empirical size of the $Wald$-test for $\tilde{\beta}_1 = 0$ in test regression (6.9) (10000 replications)[a]

Parameters			Nominal size		
T	ρ_0	ρ_1	0.10	0.05	0.01
250	0.80	0.2	0.093	0.045	0.010
250	0.25	0.75	0.081	0.037	0.006
250	0.05	0.95	0.047	0.018	0.002
1000	0.80	0.2	0.096	0.046	0.006
1000	0.25	0.75	0.096	0.047	0.009
1000	0.05	0.95	0.070	0.029	0.003

[a] The DGP is $y_t = \rho_0 + \rho_1 y_{t-1} + \nu_t$ with $\nu_t \sim NID(0, 1)$ for $t = 1, \ldots, T$.

To investigate whether the nonlinearity test has power against our smooth transition specification, we consider another simulation study. The data generating process [DGP] is given by

$$y_t = \rho_0 + \rho_1 y_{t-1} + \rho_{1,0} y_{t-1} G_0(p_{t|t-1}; \gamma, \kappa) + \nu_t \qquad \text{for} \qquad t = 1, \ldots, T, \tag{6.12}$$

where $\rho_{1,0}$ is the adjustment of the autoregressive parameter when $G_0(\cdot)$ is equal to 1. Hence, we now simulate under a specific alternative of nonlinearity.

Table 6.2 displays the power of the Wald test for $\tilde{\beta} = 0$ in the test regression (6.9) for different parameter values based on the nominal size of 5%. Results are again based on 10000 replications. We compare large and small autoregressive terms ρ_1, different distances from linearity $\rho_{1,0}$ and different parameter values for γ and κ.

Several conclusions can be drawn from the table. First of all, as expected, the power is larger for a larger sample size. Secondly, the power is larger when the alternative is further away from the null hypothesis. These are familiar aspects of the power of a statistical test. Thirdly, a larger autoregressive parameter ρ_1 leads to larger power of the test. Higher persistence in the time series leads to smaller standard errors and hence it becomes easier to detect nonlinearities. Fourth, for large γ the breaks are more prominent and easier to detect which results in larger statistical power. Finally, a threshold parameter κ which is further from the unconditional mean in the largest regime results in more separate regimes. It is therefore easier to detect the two regimes and hence the power increases.

Most importantly, the power properties of the test for our specification have a similar pattern as for the standard STAR model. Since we include an estimate of $p_{t|t-1}$ in the test regression instead of

Table 6.2: Power of the $Wald$-test for $\tilde{\beta}_0$ in test regression (6.9) for a nominal size of 5% (10000 replications)[a]

		$T = 250$				$T = 1000$			
		γ_A[b]		γ_B		γ_A		γ_B	
ρ_1	$\rho_{1,0}$	κ_A[c]	κ_B	κ_A	κ_B	κ_A	κ_B	κ_A	κ_B
	-0.05	0.048	0.050	0.047	0.048	0.050	0.053	0.052	0.049
	-0.10	0.045	0.050	0.047	0.048	0.051	0.067	0.050	0.050
0.3	-0.15	0.049	0.053	0.047	0.048	0.057	0.077	0.050	0.061
	-0.20	0.048	0.062	0.047	0.051	0.062	0.105	0.048	0.064
	-0.25	0.052	0.062	0.048	0.053	0.069	0.131	0.047	0.075
	-0.05	0.042	0.049	0.043	0.058	0.061	0.084	0.066	0.114
	-0.10	0.053	0.066	0.057	0.093	0.113	0.177	0.123	0.301
0.6	-0.15	0.066	0.098	0.077	0.157	0.189	0.305	0.212	0.548
	-0.20	0.087	0.131	0.108	0.257	0.278	0.460	0.345	0.779
	-0.25	0.117	0.176	0.145	0.381	0.369	0.582	0.492	0.923
	-0.05	0.043	0.047	0.053	0.050	0.176	0.171	0.204	0.197
	-0.10	0.089	0.095	0.107	0.114	0.441	0.458	0.550	0.542
0.9	-0.15	0.146	0.159	0.197	0.209	0.671	0.716	0.799	0.812
	-0.20	0.209	0.234	0.292	0.312	0.818	0.857	0.922	0.939
	-0.25	0.273	0.306	0.404	0.424	0.902	0.937	0.970	0.981

[a] The DGP is given in (6.12).
[b] Slow transition is obtained by putting $\gamma_A = 2.3$. The transition function covers approximately 50% of the range of the data. With $\gamma_B = 11.5$ the transition function covers 10% of the data range indicating a fast transition.
[c] Parameter κ_A equals the unconditional mean of (the largest) regime 1 plus 1 standard deviation (transition function is larger than 0.5 for about 15.9% of the data). Parameter κ_B equals the unconditional mean of regime 1 plus 1.5 standard deviation (transition function is larger than 0.5 for about 6.7% of the data).

its true value, the power is smaller than in regular STAR models. Unreported results however show that the loss in power is relatively small.

Based on the results from the two simulation studies we conclude that the adjusted version of the nonlinearity test of Luukkonen et al. (1988) can be used for our specific type of nonlinearity. For testing for an additional third regime we refer to van Dijk (1999). These nonlinearity tests are used in the next section to test for STAR type nonlinearity in US inflation data.

6.4 Application

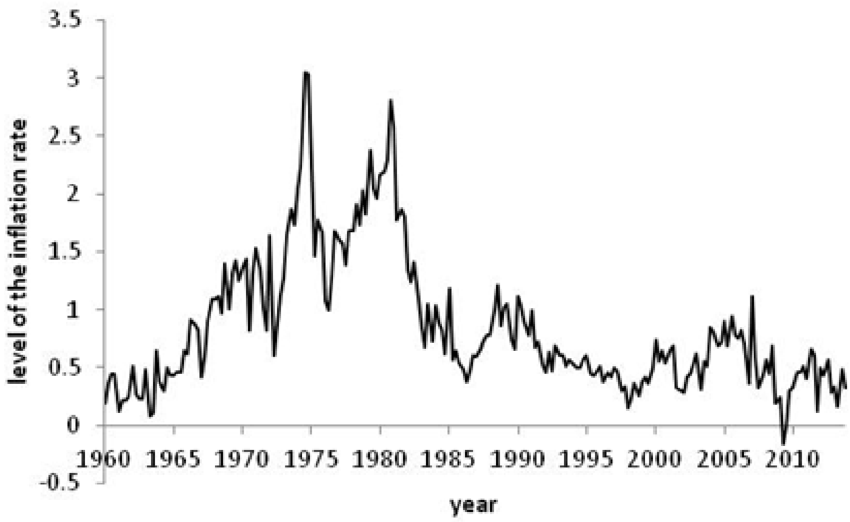
We apply the model discussed in Section 6.2 to the seasonally adjusted quarterly gross domestic product deflator based US inflation rate [henceforth called US inflation] over the period 1960.Q1-2013.Q4.[2] There are many (potentially influential) forecasts available for this inflation series (Fama and Gibbons, 1984). In this application we use the University of Michigan inflation expectation series [henceforth called Michigan series] which is a widely accepted example of an inflation forecast series created by a large number of consumers (Curtin, 1982).

In Section 6.4.1 we describe several specifications of the STAR model to describe US inflation. In Section 6.4.2 we discuss selecting the appropriate model. Section 6.4.3 deals with parameter interpretation. As the model is highly nonlinear, marginal effects (Section 6.4.3) and impulse responses (Section 6.4.3) are used.

6.4.1 Model specification

Figure 6.1 displays a plot of US inflation. It is clear from the figure that inflation peaked in the 1970s and 1980s because of the oil crises (Byrne and Davis, 2004) and became less volatile in the second half of the 1980s [great moderation,][](Rossi and Sekhposyan, 2010). The inflation rate is almost never negative in this period: deflation is only found in 2009.Q2 during the latest financial crisis.

Figure 6.1: Seasonally adjusted quarterly GDP US inflation rate 1960.Q1-2013.Q4



[2]This data is publicly available at http://www.phil.frb.org/research-and-data/ where we use the revised data series.

To model US inflation we first consider a simple linear AR model where we include an intercept and the Michigan series. There are many potential predictors of inflation (Stock and Watson, 2007; Groen et al., 2013) but this simple structure allows us to fully focus on regime changes in the inflation series itself. According to the Schwarz (1978) information criterion the appropriate lag order is 2. LM-tests indicate no serial correlation in the residuals.

As exogenous influential forecast we opt for the widely accepted Michigan series as this series is available over the whole sample period and shows a correlation of 0.88 with US inflation. We further include the endogenous forecast $p_{t|t-1}$ to the analysis and a regular STAR model with y_{t-1} as switching variable for comparison reasons.

It is clear from Figure 6.1 that a constant threshold parameter results in a model where the two oil crises are in regime 2 where inflation and hence forecasts of inflation are high. However, a large forecast in this high inflation period is different from a large forecast in the 1990s. We therefore assume that the time-varying threshold parameter will be preferred over constant thresholds. A grid search over $d = 1, \ldots, 20$ in $\bar{y}_t^{(d)}$ shows that $d = 8$ yields in general the best fit in terms of root mean squared error. This suggests that agents compare the level of the forecast to the level of US inflation in the previous two years.

6.4.2 Model selection

Before we can adopt the model specification in (6.1)-(6.2), we test for our specific form of nonlinearity. The first two rows of Table 6.3 display the results for the nonlinearity test described in Section 6.3.2. The starting point for these tests in the $ARX(2)$-specification. The first row shows that in our proposed model specification the hypothesis of linearity is clearly rejected in favor of an additional regime. This does not hold for the regular STAR specification. Further, a third regime is a significant improvement for the majority of specifications. Hence, the test results are in favor of our three-regime model specification.

The second panel of Table 6.3 shows that there is no indication for severe misspecification in the nonlinear models. Ramsey (1969) $RESET$-tests indicate that there is no neglected nonlinearity in the series. LM-tests for first and first-to-second order serial correlation in the residuals (Breusch, 1978; Godfrey, 1978) do not indicate misspecification. Tests for first and first-to-second order $ARCH$ effects do not find $ARCH$ effects in the residuals except for $ARCH(2)$ in the specification with the exogenous forecast and $\kappa_{it} = \kappa_i \sigma_t + \bar{y}_t^{(d)}$. In sum, these test results give a justification for using the model as explained in Section 6.2.

Table 6.3: Nonlinearity and misspecification tests (p-values) for the 6 model specifications for US inflation[a]

		$\kappa_t = \kappa + \bar{y}_{t-1\|t-d}$			$\kappa_t = \kappa\hat{\sigma}_t + \bar{y}_{t-1\|t-d}$		
		STAR	exogenous	endogenous	STAR	exogenous	endogenous
Nonlinearity	second regime	0.372	0.001	0.007	0.389	0.001	0.007
	third regime	0.153	0.794	0.023	0.000	0.000	0.002
RESET-test		0.080	0.041	0.777	0.336	0.011	0.196
Serial Correlation	first-order	0.548	0.946	0.063	0.722	0.935	0.917
	first-to-second order	0.554	0.896	0.071	0.486	0.801	0.990
ARCH-effects	first-order	1.000	0.597	0.645	0.458	0.226	0.394
	first-to-second order	0.790	0.380	0.806	0.211	0.004	0.115

[a] The tests are the adjusted nonlinearity test by Luukkonen et al. (1988), the Ramsey (1969) RESET test, the serial correlation test by Breusch (1978); Godfrey (1978) and the ARCH LM-test for heteroskedasticity by Engle (1982).

For model selection purposes, we compare the fit of the model specifications, that is the regular STAR, the exogenous and the endogenous models for both threshold specifications. Since these models are non-nested, standard likelihood ratio tests cannot be used. We therefore opt for the Vuong (1989) test based on the assumption of normality of the disturbances. Further, we use a nonparametric sign test on the absolute value of the residuals (Dixon and Mood, 1946).

Table 6.4 displays the test statistics. The Vuong (1989) test does not favor the nonlinear model specifications compared to the $ARX(2)$-model because of the large increase of the number of parameters. The nonlinearity test in Table 6.3, which is especially designed to compare linear AR models versus STAR models however indicated that adding nonlinearity leads to improvements of the model. The nonparametric sign test shows more support for this claim concerning the $ARX(2)$-specification. Since we want to choose one specification, we opt for the model with the lowest sum of squared residuals [SSR].

We compare the in-sample endogenous forecast of the selected model specification with the exogenous Michigan series and y_{t-1}. The bias of the forecast is 0.176 compared to 0.349 for the Michigan series and 0.195 for y_{t-1}. The root mean squared prediction error is 0.226 versus 0.409 and 0.267 for the endogenous, exogenous and random walk forecasts, respectively. Again, these results support our endogenous forecast specification in (6.3).

In sum, based on the test results and model fit we opt for the nonlinear specification with the endogenous forecast and influential forecast $p_{t|t-1}$ and $\kappa_{it} = \kappa_i + \bar{y}_t^{(d)}$.

Table 6.4: Vuong (1989) and sign tests for comparing the 6 different specifications and an ARX(2)-model for US inflation (p-values in parentheses)[a]

		ARX(2)	$\kappa_t = \kappa + \bar{y}_{t-1\|t-d}$			$\kappa_t = \kappa\hat{\sigma}_t + \bar{y}_{t-1\|t-d}$		
			STAR[b]	Exo	Endo	STAR	Exo	Endo
SSR		1[c]	0.907	0.849	0.806	0.884	0.844	0.813
ARX(2)			2.954	2.335	1.539	5.672	2.564	1.054
			(0.003)	(0.020)	(0.124)	(0.000)	(0.010)	(0.292)
$\kappa_t = \kappa + \bar{y}_{t-1\|t-d}$	STAR	0.440		-0.762	-0.888	0.643	-0.608	-1.665
		(0.044)		(0.446)	(0.375)	(0.520)	(0.543)	(0.096)
	Exo	0.449	0.454		-0.210	1.442	0.572	-0.674
		(0.076)	(0.098)		(0.833)	(0.149)	(0.567)	(0.500)
	Endo	0.440	0.477	0.495		1.438	0.348	-0.371
		(0.044)	(0.270)	(0.473)		(0.150)	(0.728)	(0.711)
$\kappa_t = \kappa\hat{\sigma}_t + \bar{y}_{t-1\|t-d}$	STAR	0.514	0.519	0.491	0.500		-1.273	-1.683
		(0.317)	(0.270)	(0.419)	(0.473)		(0.203)	(0.092)
	Exo	0.486	0.519	0.528	0.528	0.509		-0.770
		(0.367)	(0.270)	(0.188)	(0.188)	(0.367)		(0.441)
	Endo	0.486	0.481	0.514	0.514	0.500	0.468	
		(0.367)	(0.317)	(0.317)	(0.317)	(0.473)	(0.188)	

[a] The upper-triangular matrix in the table shows the results for the Vuong (1989) test. A positive test value indicates that the model presented in the row is better than the model in the column. The lower-triangular matrix displays the sign test. A test value smaller than 0.5 indicates that the model presented in the row is better.

[b] 'STAR' stands for the regular STAR-model, 'Exo' stands for the model with the exogenous forecast, 'Endo' stands for the model with the endogenous forecast $p_{t|t-1}$

[c] The sum of squared residuals [SSR] for the ARX(2) specification is normalized to 1.

6.4.3 Parameter interpretation

Table 6.5 and Tables 6.A.1 and 6.A.2 display the parameter estimates of the model specifications. We focus on the interpretation of the estimation results of the specification selected in Section 6.4.2 which is shown in the third panel of Table 6.5. At first glance we see that the estimates for γ_i are relatively small indicating a smooth transition from regime to regime.

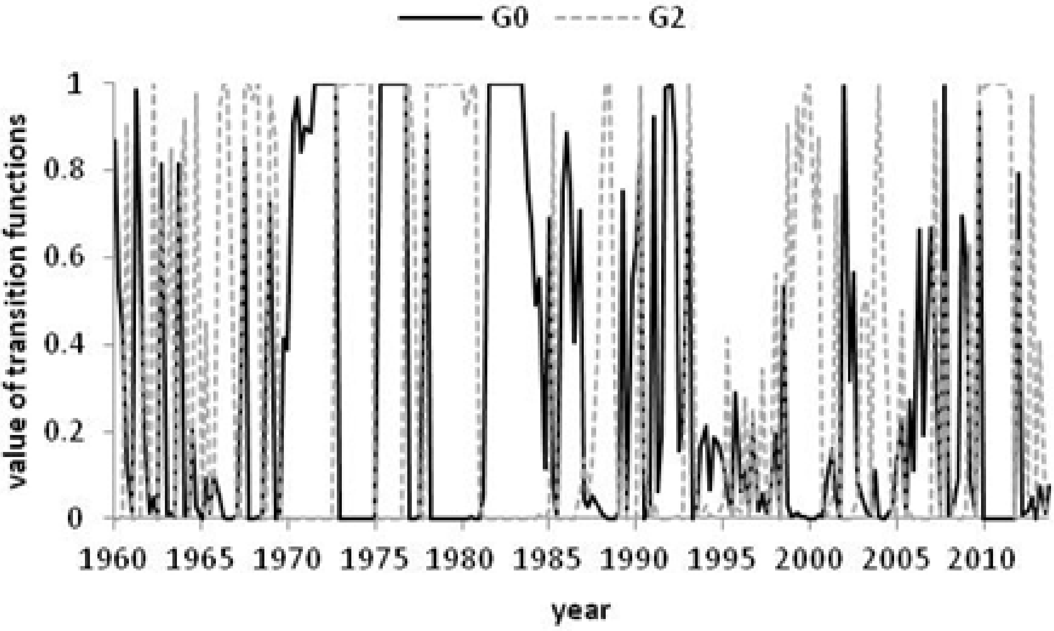
As the structure of the model is highly nonlinear direct interpretations of individual parameter estimates is difficult. We therefore consider several graphs to investigate features of US inflation and the impact of forecasts. Figure 6.2 plots the values of the transition functions over time. The spikes in the transition function for regime 2 during the oil crises show that the model can distinguish high from moderate forecasts during these crises. Since $\bar{y}_t^{(d)}$ is relatively large just after these crises, the low forecast regime dominates. Inflation targeting led to a steady US inflation rate in the 1990s and hence, the inflation rate has mostly been in the intermediate regime.

Table 6.5: WNLS parameter estimates of the 3 model specifications with $\kappa_{i,t} = \kappa + \bar{y}_{t-1|t-d}$ for US inflation (standard errors in parentheses)

	regime 0[a]		regime 1		regime 2[a]	
STAR						
κ	-0.363	(0.328)			0.387	(10.851)
γ	-293.377	–			293.379	–
c	0.153	(0.391)	-0.140	(0.044)	-0.437	(0.327)
$INFL_{t-1}$	-0.099	(0.269)	0.451	(0.083)	-0.300	(0.573)
$INFL_{t-2}$	-0.195	(0.265)	0.169	(0.069)	0.275	(0.481)
MS_{t-1}	0.245	(0.475)	0.381	(0.060)	0.374	(0.224)
κ_{gm}			1985.Q3	(4.779)		
γ_{gm}			5.882	–		
σ_1^2			0.057	(0.008)		
$\sigma_2^2 - \sigma_1^2$			-0.028	(0.009)		
Exogenous						
κ	0.064	(0.015)			0.370	(0.017)
γ	-246.640	–			246.659	–
c	0.145	(0.143)	-0.189	(0.094)	0.058	(0.121)
$INFL_{t-1}$	0.268	(0.158)	0.113	(0.108)	0.540	(0.169)
$INFL_{t-2}$	-0.348	(0.151)	0.303	(0.105)	-0.070	(0.168)
MS_{t-1}	0.036	(0.225)	0.608	(0.174)	-0.398	(0.200)
κ_{gm}			1984.Q3	(18.238)		
γ_{gm}			0.171	–		
σ_1^2			0.054	(0.008)		
$\sigma_2^2 - \sigma_1^2$			-0.026	(0.009)		
Endogenous						
κ	-0.067	(0.022)			0.081	(0.016)
γ	-30.628	–			56.547	–
c	-0.500	(0.084)	0.296	(0.061)	-0.475	(0.061)
$INFL_{t-1}$	0.466	(0.083)	-0.170	(0.059)	0.834	(0.055)
$INFL_{t-2}$	-0.699	(0.096)	0.760	(0.089)	-0.801	(0.090)
MS_{t-1}	0.826	(0.115)	-0.082	(0.091)	0.505	(0.105)
κ_{gm}			1981.Q1	(3.644)		
γ_{gm}			5.882	–		
σ_1^2			0.058	(0.007)		
$\sigma_2^2 - \sigma_1^2$			-0.032	(0.008)		

[a] Parameters of regime 0 and regime 2 are in difference with regime 1

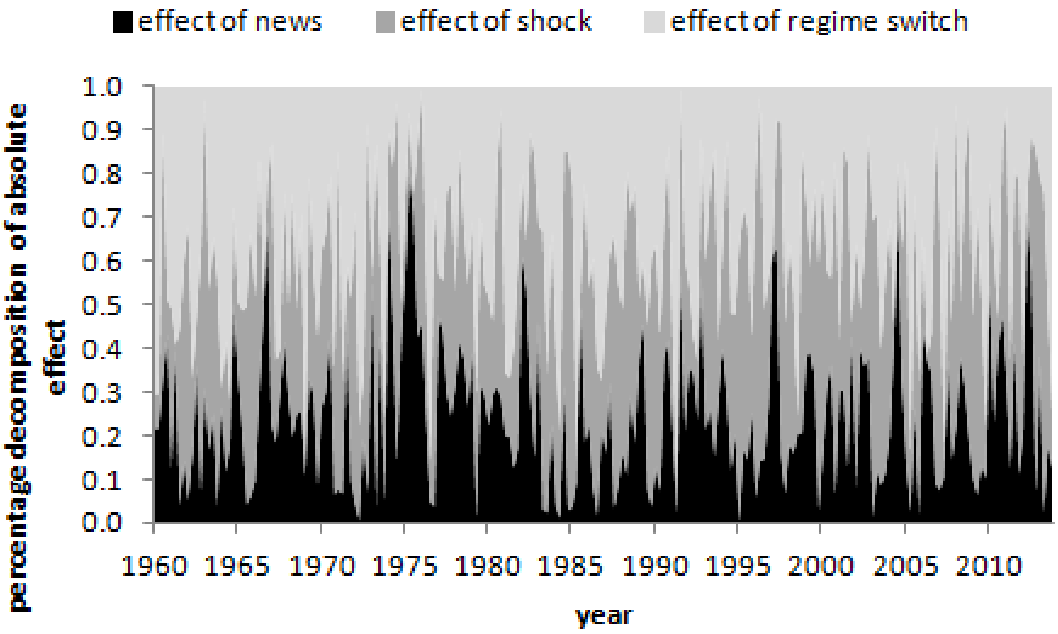
Figure 6.2: Transition functions for US inflation for the model with an endogenous forecast $p_{t|t-1}$ and a time-varying threshold parameter $\kappa_{i,t} = \kappa_i + \bar{y}_{t-1|t-d}$



The parameter estimates of the time-varying variance are also shown in Table 6.5. These imply that a decrease in variance in the chosen model took place in the first quarter of 1981. This date is somewhat earlier than reported in other studies (Kahn et al., 2002) and earlier than in the other specifications (1985.Q3).

To shed more light on the effect of forecasts we decompose the change in US inflation in three parts. The first part concerns the effect of changes in explanatory variables holding the regime constant (average absolute effect of 0.11). The second part is the error term (average absolute effect of 0.16). The third part describes the effect of the forecast-based regime switches (average absolute effect of 0.17). Figure 6.3 displays the decomposition over time in percentage points. The effect of the forecast-based regime switches is largest in about 45% of the time.

Figure 6.3: Percentage decomposition of the absolute effect of changes over time ($y_t - y_{t-1}$) in US inflation for the model with an endogenous forecast $p_{t|t-1}$ and a time-varying threshold parameter $\kappa_{i,t} = \kappa_i + \bar{y}_{t-1|t-d}$



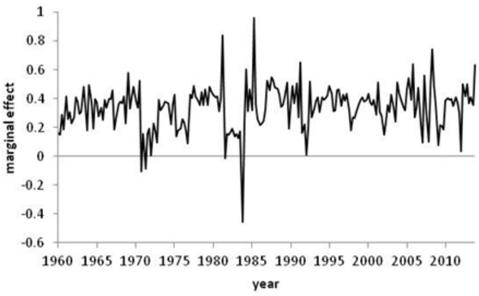
Marginal effects

To analyze the differences in dynamic patterns across the three regimes, we consider marginal effects. Marginal effects are defined as the change in y caused by one standard deviation increase in x, where x denotes lagged values of US inflation or the Michigan series. Note that the change in x can also cause regime switches to occur. Marginal effects therefore differ over time and are plotted in Figure 6.4. Further, Table 6.6 displays the average marginal effects weighted with the regimes.

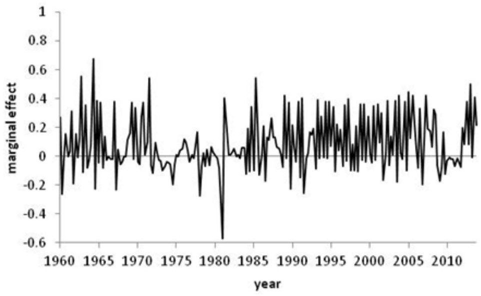
Table 6.6 shows that the first lag of inflation on average has a larger impact on the inflation rate in the higher regimes. This indicates that agents rely more on the near past in periods with relatively high forecasts. The second lag of inflation has a smaller absolute impact in the outer regimes. Thus, the distant past is less important to agents in economically uncertain periods. Further, the influence of the Michigan series is similar in all regimes. This partly contradicts the findings of Carroll (2003). Where they find that in periods of extreme forecasts people tend to adjust their personal forecasts to the expert opinion, we find that actions to adjust the inflation rate are as much based on the widely accepted Michigan series as in the intermediate regime.

The last panel of Figure 6.4 shows the marginal effect of a positive change in $p_{t|t-1}$. For all regimes this effect is on average negative. That is, an increase in the forecast adjusts the inflation rate downward. This effect is smallest in the intermediate and economically stable regime. If regime 0 is very prominent, the marginal effect is large and positive. Hence, for relatively small forecasts the adjustments are upward. Agents thus behave such that the inflation rate is mean reverting: an increase in the forecast leads to reactions which lower the inflation rate if the forecast is relatively large, but increase the inflation rate if the forecast was still relatively low. This contradicts the expectations trap literature (Christiano and Gust, 2000), where an upward change in inflation is expected when forecasts are large. It is nonetheless in line with inflation targeting since the FED Volcker regime: deviation from the target inflation rate leads to reversing actions.

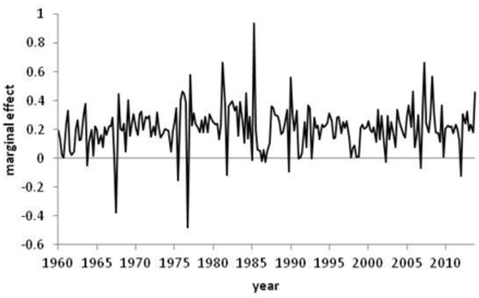
Figure 6.4: Marginal effects in the application of US inflation of a one standard deviation increase in the explanatory variables and $p_{t|t-1}$ for the model with an endogenous forecast $p_{t|t-1}$ and a time-varying threshold parameter $\kappa_{i,t} = \kappa_i + \bar{y}_{t-1|t-d}$



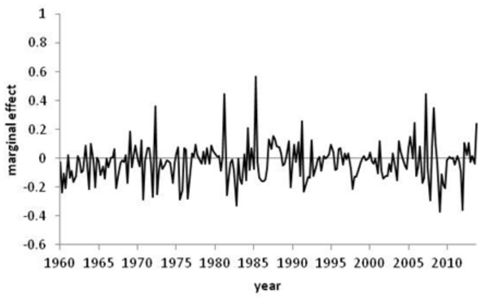
(a) marginal effect of $INFL_{t-1}$



(b) marginal effect of $INFL_{t-2}$



(c) marginal effect of MS_{t-1}



(d) marginal effect of $p_{t|t-1}$

Table 6.6: Descriptive statistics of marginal effects as displayed in Figure 6.4[a]

	y_{t-1}			y_{t-2}			ms_{t-1}			$p_{t\|t-1}$		
	5%[b]	average	95%	5%[b]	average	95%	5%	average	95%	5%	average	95%
	0.098	0.357	0.537	-0.186	0.023	0.411	-0.016	0.215	0.445	-0.237	-0.022	0.152
regime 0	-0.009	0.193	0.522	-0.186	0.026	0.381	-0.120	0.216	0.445	-0.290	-0.090	0.152
regime 1	0.183	0.392	0.639	-0.194	0.138	0.423	0.000	0.216	0.463	-0.209	-0.002	0.254
regime 2	0.282	0.383	0.484	-0.166	-0.013	0.394	0.123	0.214	0.365	-0.125	-0.011	0.090

[a] The first row shows the equally weighted marginal effects. The second to fourth rows show the weighted marginal effect where the weights are based on the probability to be in the specific regime.

[b] 5% stands for the 5 percent percentile, while 95% stands for the 95 percent percentile.

Impulse response analysis

To interpret the dynamic properties of the model we use generalized impulse response functions (Koop et al., 1996, GIRF). We examine the impact of a shock δ for different information sets Ω_t in a similar way as in van Dijk (1999). The GIRF is defined as

$$GIRF_y(h, \delta, \Omega_\tau) = E[y_{\tau+h}|\varepsilon_\tau = \varepsilon_\tau + \delta, \Omega_\tau] - E[y_{\tau+h}|\Omega_\tau], \tag{6.13}$$

where τ denotes the timing of the shock, h is the horizon and where Ω_τ is the information set at time τ. Hence, the impulse response function denotes the dynamic effect of a shock δ at time τ on future values of y_t. We average over all possible information sets Ω_τ and we also split the results depending on the regime at time τ. Note that a shock may also affect future regimes and the analysis thus takes full advantage of the nonlinearity of the model specification.

We further define the π-absorption time of the shock as the amount of time periods it takes before $\pi\%$ of the shock is absorbed (van Dijk et al., 2007), that is

$$A_y(\pi, \delta, \Omega_t) = \sum_{m=0}^{\infty}(1 - \prod_{h=m}^{\infty} I_y(\pi, h, \delta, \Omega_t)), \tag{6.14}$$

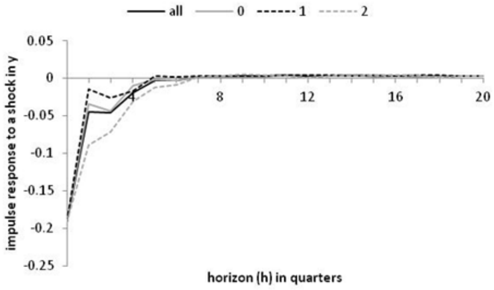
where

$$I_y(\pi, h, \delta, \Omega_t) = I[|GIRF_y(h, \delta, \Omega_t)| \leq \pi|\delta|] \tag{6.15}$$

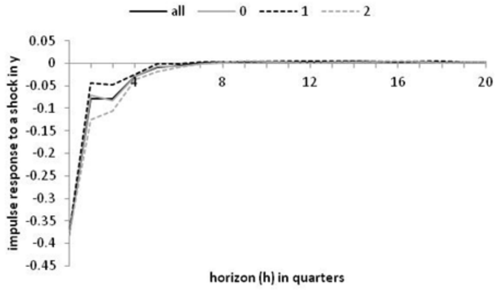
with $I[\cdot]$ an indicator function which is 1 if the argument is true and 0 otherwise.

Figure 6.5 displays the impulse response function for positive and negative shocks in y_τ for different regimes. The differences across regimes are relatively small although the reaction to a shock in regime 2 has a larger absorption time. Only for regime 2 it takes more than one quarter to absorb 50% of the shock. Nonetheless, it takes for all shocks 3 to 6 quarters until 90% of the shock is absorbed. Hence, an innovative shock has a small but relatively long-lasting effect on the inflation rate in the future.

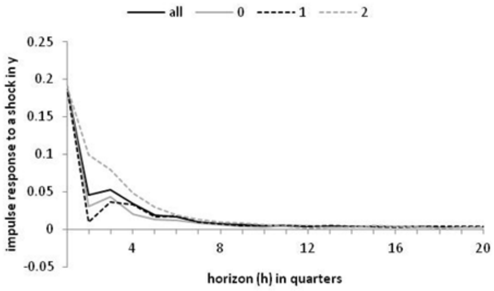
Figure 6.5: Impulse response analysis in the application of US inflation for various shocks δ to y_t for the model with an endogenous forecast $p_{t|t-1}$ and a time-varying threshold parameter $\kappa_{i,t} = \kappa_i + \bar{y}_{t-1|t-d}$



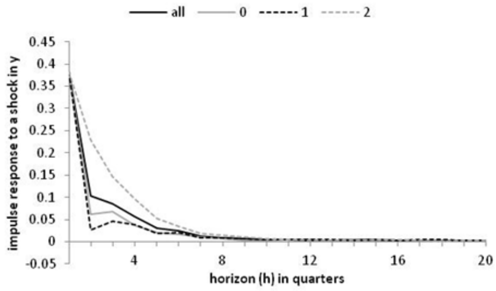
(a) $\delta = -\hat{\sigma}_\tau$



(b) $\delta = -2\hat{\sigma}_\tau$



(c) $\delta = \hat{\sigma}_\tau$



(d) $\delta = 2\hat{\sigma}_\tau$

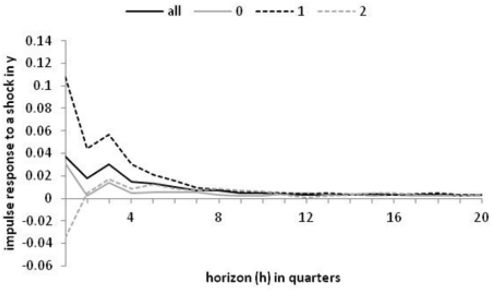
Given the structure of the model it is perhaps more interesting to examine the effect of a shock to the forecast $p_{t|t-1}$

$$GIRF_p(h, \delta, \Omega_\tau) = E[y_{\tau+h}|\Omega_\tau, p_{\tau|\tau-1} = p_{\tau|\tau-1} + \delta] - E[y_{\tau+h}|\Omega_\tau]. \tag{6.16}$$

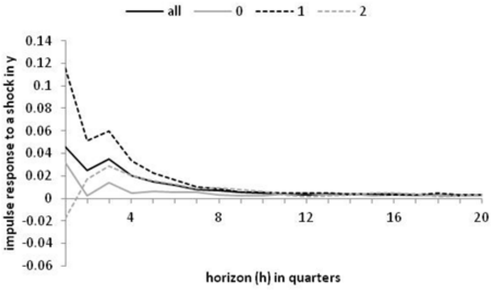
Figure 6.6 shows the effect of a shock to the forecast $p_{t|t-1}$ for shocks of various magnitudes and for different regimes at time τ.[3] The first two graphs show that a negative shock has a positive impact on the future inflation rate in regime 0 and 1. That is, a decrease in the forecast $p_{t|t-1}$ leads to an increase in the inflation rate for approximately a year. Hence, agents try to correct for a predicted decrease when inflation is already in the lower regimes. The effect of a negative shock in regime 2 starts off negative, thus indicating mean reverting behavior of economic agents. The bottom two graphs indicate that a positive shock results in immediate adjusting behavior in regime 0, which is corrected in the long run. These effects of shocks to the forecast are small but again long-lasting. For example, it takes on average 5 quarters before a negative shock of magnitude σ_τ in regime 1 is absorbed for more than 90%.

[3]Note that this is a theoretical exercise as the model does not explicitly allow for a random shock to the forecast.

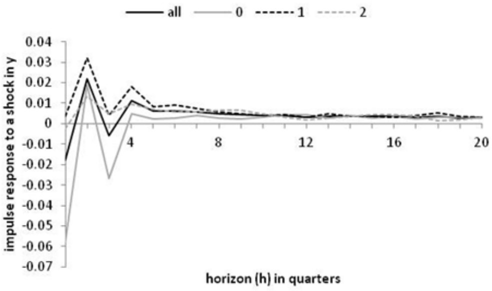
Figure 6.6: Impulse response analysis in the application of US inflation for various shocks δ to $p_{t|t-1}$ for the model with an endogenous forecast $p_{t|t-1}$ and a time-varying threshold parameter $\kappa_{i,t} = \kappa_i + \bar{y}_{t-1|t-d}$



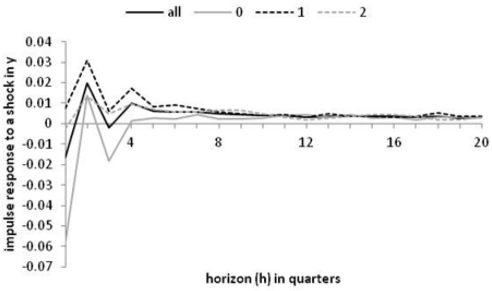
(a) $\delta = -\hat{\sigma}_\tau$



(b) $\delta = -2\hat{\sigma}_\tau$



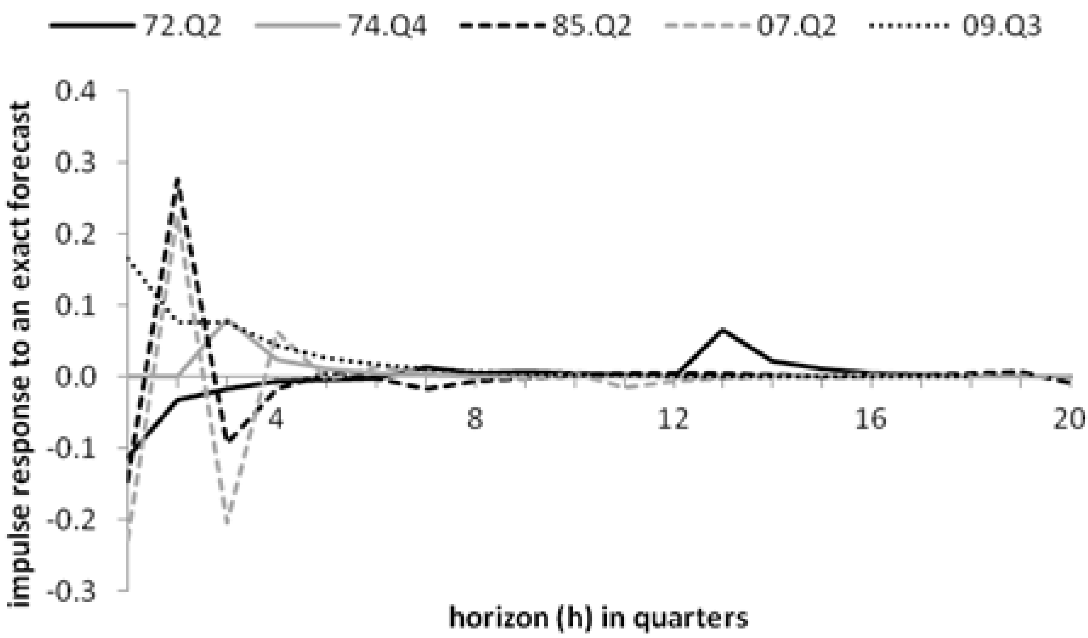
(c) $\delta = \hat{\sigma}_\tau$



(d) $\delta = 2\hat{\sigma}_\tau$

We finally consider the hypothetical situation where we impose a shock to the forecast which makes the forecast equal to the realization. This analysis investigates the importance of forecast accuracy and whether extreme events could have been circumvented with more accurate forecasts. Figure 6.7 displays impulse response functions for five data points where the forecast $p_{t|t-1}$ was inaccurate. For example, the financial crisis was at its peak in the third quarter of 2009 and the forecast has not been capable of capturing this downward spike in inflation. If the forecast $p_{t|t-1}$ had been correct, the inflation rate would have been higher for approximately 1.5 years bringing it towards the level of 2005. Further, reactions to the sudden peaks in 1985.Q1 and 2007.Q1 are similar and more accurate forecasts would have changed the inflation rate for about a year. Finally, if the peak in inflation in the oil crisis in 1974.Q3 would have been forecasted correctly, the trough in Figure 6.1 had been smoother. This hypothetical analysis shows the importance of accurate forecasts: where forecasts have been inaccurate a more accurate forecast would have highly changed agents reactions and thereby the future inflation rate.

Figure 6.7: Impulse response analysis in the application of US inflation of a shock in $p_{\tau|\tau-1}$ which makes the forecast exactly equal to the dependent variable. Five quarters where the forecast of inflation is far from the realization are displayed.



In sum, we find that our model proposed in Section 6.2 is capable of capturing familiar aspects of US inflation. Marginal effects and impulse response analyses show that agents take the forecast of the dependent variable into account when they take action at the economic market. The model especially shows that relatively low forecasts result in structural reactions of agents which cause the inflation rate to be higher than the original forecast.

6.5 Conclusion

In this chapter we introduced a novel STAR type time series model where regime switched are based on the relative size of the forecast of the underlying time series. The forecast determining regime switches can either be exogenous to the model or based on a forecast from the model itself. The specification analyzes the impact of forecasts based on whether the forecast is relatively high or low. The time series model describes macro-economic variables where it is likely that forecasts have an impact on the dependent variable.

The model is applied to US GDP deflator inflation rate. Since the level of inflation changes over time we include a time-varying threshold parameter in the L-STAR specification such that the relative size of the forecast determines regime changes.

Empirical results show that (i) forecasts lead to regime changes and have an impact on the level of inflation; (ii) a relatively large forecast results in actions which lower the inflation rate; (iii) the absorption time of negative shocks in the forecast of inflation is relatively large and the effect of these shocks is positive in the long run, indicating mean reverting behavior and; (iv) a counter-factual scenario where forecasts during the financial crisis in 20009 where assumed to be correct would have resulted in a higher level of inflation in the subsequent quarters.

The model in this chapter is applicable to (macro-economic) variables which are likely to be prone to forecasts. The impact of forecasts in other key variables is of future interest. Several other extensions are possible. For example, we now assume that the reaction to one-step ahead forecasts already take place in the next quarter. Nevertheless, reactions of agents may be slow. Hence, the forecast of today may lead to regime changes in later quarters. Another extension may be to consider a Philips curve type of model and allow the effect of predictors to change according to the relative size of the forecast.

6.A Tables

Table 6.A.1: WNLS parameter estimates of the 3 model specifications with $\kappa_{i,t} = \kappa$ for US inflation (standard errors in parentheses)

	regime 0[a]		regime 1		regime 2[a]	
STAR						
κ	0.603	(0.017)			1.227	(0.200)
γ	-198.625	–			198.612	–
c	0.341	(0.188)	-0.251	(0.152)	0.213	(0.213)
$INFL_{t-1}$	0.280	(0.252)	0.108	(0.182)	0.349	(0.223)
$INFL_{t-2}$	-0.109	(0.165)	0.289	(0.120)	-0.299	(0.164)
MS_{t-1}	-0.519	(0.202)	0.657	(0.157)	-0.130	(0.177)
κ_{gm}			1985.Q4	(3.737)		
γ_{gm}			5.880	–		
σ_1^2			0.057	(0.007)		
$\sigma_2^2 - \sigma_1^2$			-0.029	(0.007)		
Exogenous						
κ	0.475	(0.743)			1.261	(0.084)
γ	-4.968	–			227.138	–
c	0.661	(3.114)	0.293	(0.824)	-0.313	(0.743)
$INFL_{t-1}$	-0.860	(2.638)	0.407	(0.149)	0.192	(0.186)
$INFL_{t-2}$	-1.076	(2.524)	0.369	(0.161)	-0.350	(0.193)
MS_{t-1}	-0.859	(4.525)	-0.066	(0.704)	0.431	(0.667)
κ_{gm}			1985.Q3	(7.894)		
γ_{gm}			5.318	–		
σ_1^2			0.058	(0.007)		
$\sigma_2^2 - \sigma_1^2$			-0.028	(0.008)		
Endogenous						
κ	0.603	(0.031)			0.791	(0.025)
γ	-175.689	–			120.575	–
c	-0.676	(0.186)	0.772	(0.172)	-0.942	(0.166)
$INFL_{t-1}$	0.782	(0.235)	-0.455	(0.191)	1.040	(0.182)
$INFL_{t-2}$	0.252	(0.182)	-0.064	(0.113)	0.027	(0.111)
MS_{t-1}	-0.352	(0.266)	0.496	(0.255)	-0.003	(0.240)
κ_{gm}			1981.Q3	(2.786)		
γ_{gm}			5.851	–		
σ_1^2			0.063	(0.008)		
$\sigma_2^2 - \sigma_1^2$			-0.039	(0.008)		

[a] Parameters of regime 0 and regime 2 are in difference with regime 1

Table 6.A.2: WNLS parameter estimates of the 3 model specifications with $\kappa_{i,t} = \kappa\hat{\sigma}_t + \bar{y}_{t-1|t-d}$ for US inflation (standard errors in parentheses)

	regime 0[a]		regime 1		regime 2[a]	
STAR						
κ	-0.484	(1.360)			0.397	(0.870)
γ	-2.719	–			2.717	–
c	-1.143	(1.796)	0.588	(2.099)	-1.855	(1.421)
$INFL_{t-1}$	-0.838	(1.320)	0.842	(1.242)	-0.840	(1.248)
$INFL_{t-2}$	0.849	(1.273)	-0.313	(1.352)	0.969	(1.263)
MS_{t-1}	0.578	(1.200)	0.078	(1.063)	0.962	(0.804)
κ_{gm}			1981.Q3	(2.580)		
γ_{gm}			5.885	–		
σ_1^2			0.064	(0.009)		
$\sigma_2^2 - \sigma_1^2$			-0.034	(0.010)		
Exogenous						
κ	0.054	(0.017)			0.443	(0.037)
γ	-246.662	–			246.609	–
c	0.231	(0.135)	-0.255	(0.074)	0.117	(0.108)
$INFL_{t-1}$	0.190	(0.155)	0.170	(0.103)	0.467	(0.167)
$INFL_{t-2}$	-0.270	(0.150)	0.237	(0.099)	-0.020	(0.170)
MS_{t-1}	-0.050	(0.213)	0.692	(0.157)	-0.466	(0.189)
κ_{gm}			1985.Q3	(24.862)		
γ_{gm}			5.879	–		
σ_1^2			0.054	(0.008)		
$\sigma_2^2 - \sigma_1^2$			-0.026	(0.008)		
Endogenous						
κ	-0.292	(0.028)			0.398	(0.283)
γ	-209.647	–			293.380	–
c	0.168	(0.144)	-0.113	(0.041)	0.301	(0.741)
$INFL_{t-1}$	-0.557	(0.096)	0.460	(0.067)	0.125	(0.837)
$INFL_{t-2}$	-0.434	(0.129)	0.236	(0.059)	-0.298	(0.861)
MS_{t-1}	0.877	(0.053)	0.300	(0.054)	0.159	(0.391)
κ_{gm}			1985.Q3	(4.736)		
γ_{gm}			5.868	–		
σ_1^2			0.053	(0.008)		
$\sigma_2^2 - \sigma_1^2$			-0.025	(0.009)		

[a] Parameters of regime 0 and regime 2 are in difference with regime 1

Nederlandse Samenvatting (Summary in Dutch)

Dit proefschrift bestaat uit twee delen op het gebied van econometrisch modelleren. Allereerst onderzoek ik het schatten van parameters in (multivariate) discrete keuzemodellen. Ik behandel model complicaties wanneer de specificatie complexer wordt. Keuzemodellen geven inzicht in beslissingen van personen. Daarnaast geven ze weer wat de invloed is van andere variabelen op het keuzeproces. De nieuwe uitbreidingen op standaard keuzemodellen in dit proefschrift laten onder andere zien hoe verschillende keuzes elkaar beïnvloeden. Voor deze complexe model specificaties stel ik nieuwe methodes voor om de parameters te schatten. Als tweede behandel ik het effect van voorspellingen op macro-economische variabelen. Omdat actoren op de economische markt op belangrijke voorspellingen reageren, zullen deze voorspellingen invloed hebben op het verloop van de economie.

Discrete Keuzemodellen

Iedereen maakt iedere dag veel discrete keuzes. De reis naar werk kan bijvoorbeeld worden afgelegd per fiets, auto of openbaar vervoer. De keuze voor producten in een supermarkt is ook een voorbeeld van zo'n discreet keuzeproces. Keuzemodellen zijn bedoeld om deze keuzes te beschrijven en te zien welke aspecten de keuzes beïnvloeden. De keuze voor een bepaald merk zal bijvoorbeeld samenhangen met de prijs van het product, maar ook met het inkomen van de persoon of met advertenties.

Met behulp van econometrie kunnen relaties tussen variabelen en keuzes worden onderzocht. Econometrische keuzemodellen dateren uit omstreeks 1860 en Nobelprijs-winnaar Daniel McFadden heeft belangrijk werk geleverd. Vier hoofdstukken in dit proefschrift dragen bij aan de kennis over keuzemodellen.

Drie hoofdstukken introduceren een multivariate uitbreiding op bestaande univariate keuzemodellen. Deze hoofdstukken onderzoeken op welke manier keuze A beïnvloed wordt door keuzes B, C, et cetera. Omdat discrete keuzemodellen niet gemakkelijk uit te breiden zijn naar meer dimensies, is de literatuur over multivariate keuzes schaars. Dit proefschrift introduceert bruikbare methodes voor het vinden van parameterschattingen voor dit type model.

De multivariate extensie van keuzemodellen wordt voorgesteld voor drie verschillende situaties. Allereerst kijk ik naar zogeheten gecorreleerde binaire keuzes. Dit hoofdstuk is gebaseerd op Bel et al. (2014). Bijvoorbeeld kan gedacht worden aan de keuzes om winkels wel of niet te bezoeken. De keuze om een tweedehands winkel te bezoeken is waarschijnlijk (positief of negatief) gecorreleerd met keuzes om andere winkels in te gaan. Hiervoor gebruik ik een bestaande model specificatie zoals geïntroduceerd door Cox (1972). Een volgende stap is de uitbreiding naar meerdere multinomiale keuze die evengoed gecorreleerd kunnen zijn. Dit onderzoek is gebaseerd op Bel and Paap (2014). Het is bijvoorbeeld erg aannemelijk dat de keuze voor een goedkoop merk in productklasse A, ook een keuze voor een goedkoop merk in productklasse B oplevert. Voor deze gecorreleerde multinomiale keuzes introduceer ik een nieuwe model specificatie welke ook gemakkelijk toepasbaar is wanneer het probleem een groot aantal keuzes en keuze-opties bevat. Als laatste multivariate extensie kijk ik naar gecorreleerde geordende keuzes, gebaseerd op Bel and Schoonees (2015). Hierbij breid ik het bivariate Dale (1986) model uit naar een multivariate setting.

Omdat de voorgenoemde model specificaties gemakkelijk zeer groot worden met veel keuzes en veel keuze mogelijkheden, kan de standaard schattingsmethode maximum likelihood lastig worden gebruikt. Zo worden kansen numeriek klein en wordt de computertijd voor het schatten van de parameters erg groot. Bij problemen op grote schaal zal maximum likelihood niet in staat zijn een oplossing te geven. Daarom stel ik in het eerste hoofdstuk verschillende alternatieve schattingsmethoden voor, waarvan de composite likelihood methode het beste alternatief blijkt en in de overige hoofdstukken ook wordt gebruikt. Lindsay (1988) heeft composite likelihood geïntroduceerd als "een likelihood-type object welke gevormd wordt door individuele log-likelihood componenten samen te nemen". Het maximaliseren van deze composite log-likelihood is een consistente schattingsmethode (Varin et al., 2011), maar omdat de methode niet de exacte likelihood functie gebruikt is deze inefficiënt. De hoofdstukken in dit proefschrift laten aan de hand van simulaties zien dat het verlies van efficiëntie zeer klein is. Daarom is de reductie van complexiteit dankzij composite likelihood een duidelijk voordeel. Zelfs in het geval van grote dimensies is composite likelihood nog bij machte accurate parameter schattingen te produceren - dit in tegenstelling tot de standaard maximum likelihood methode. Dit proefschrift is het eerste om deze methode toe te passen op multivariate keuzemodellen.

Een laatste hoofdstuk over keuzemodellen bediscussieerd het keuzeproces van individuen. Dit hoofdstuk is gebaseerd op Bel and Paap (2015). Bestaande literatuur toont aan dat een multinomiale keuze vaak in twee stappen wordt gemaakt: allereerst wordt een set van overwegingen gevormd waarna hieruit een keuze wordt gemaakt. In een supermarkt wordt bijvoorbeeld niet gekozen uit alle merken, maar heeft de klant al eerder een voorselectie gemaakt. De voorselectie van de klant is vaak niet geobserveerd door de econometrist.

Ik voeg een kritische noot toe aan de literatuur. Het is namelijk moeilijk vast te stellen welke verklarende variabele welke stap van het keuzeproces beïnvloedt. Heeft de prijs van een product bijvoorbeeld effect op overweging van een product, op de uiteindelijke keuze of op allebei? Ik bekijk of de rol van de verklarende variabelen kan worden geïdentificeerd aan de hand van statistische maatstaven. De uitkomst laat zien dat dit niet goed mogelijk is en dat de schatting van de niet geobserveerde set van overwegingen alleen geïnterpreteerd kan worden als de indeling van variabelen met zekerheid overeenkomt met het data genererende proces.

De bijdrage van deze 4 hoofdstukken aan de literatuur zijn: (i) de extensie naar multivariate analyse van discrete keuzemodellen; (ii) een computationeel bruikbare schattingsmethode voor parameters in deze multivariate keuzemodellen en (iii) een kritische bijdrage aan de rol van verklarende variabelen in een twee-staps keuzeproces.

Modelleren van de Invloed van Voorspellingen

Het tweede deel van het proefschrift omvat één hoofdstuk en is gebaseerd op Bel and Paap (2013). Dit hoofdstuk beschrijft het effect van voorspellingen op macro-economische variabelen zoals inflatie. Omdat actoren op de economische markt beslissingen baseren op nieuws en voorspellingen, ga ik uit van invloed van voorspellingen op de economie. Reacties op een voorspelling van het Nederlandse Centraal Planbureau zullen bijvoorbeeld het verloop van economische variabelen veranderen.

De bijdrage van dit hoofdstuk is een nieuwe model specificatie die rekening houdt met de reactie op voorspellingen. Een smooth transition autoregressive model (Teräsvirta and Anderson, 1992) wordt gebruikt om om te gaan met de discontinuïteit van de tijdreeks veroorzaakt door de voorspelling. Deze voorspelling kan of een exogene mening van een expert zijn of door het model zelf worden gegenereerd. Voor dit laatste gebruik ik het zogenoemde contemporaneous smooth transition autoregressive model (Dueker et al., 2007).

Een applicatie van dit model op inflatie in de Verenigde Staten laat zien dat voorspellingen inderdaad invloed hebben op het niveau van inflatie. Volgens de uitkomsten van het model zorgen reacties op voorspellingen dat het niveau van inflatie weer terugkeert naar het gemiddelde ten opzichte van de extreme voorspelling.

Bibliography

Agresti, A., 2007. An introduction to categorical data analysis. Wiley-Interscience.

Albanesi, S., Chari, V. V., Christiano, L. J., 2003. Expectation traps and monetary policy. The Review of Economic Studies 70 (245), 715–741.

Altonji, J. G., Segal, L. M., 1996. Small-sample bias in gmm estimation of covariance structures. Journal of Business and Economic Statistics 14 (3), 353–366.

Amemiya, T., 1978. On a two-step estimation of a multivariate logit model. Journal of Econometrics 8 (1), 13 – 21.

Amemiya, T., 1985. Advanced Econometrics. Basil Blackwell, Oxford.

Anderson, J. A., 1984. Regression and ordered categorical variables. Journal of the Royal Statistical Society. Series B (Methodological) 46 (1), 1–30.

Andrews, R. L., Srinivasan, T. C., 1995. Studying consideration effects in empirical choice models using scanner panel data. Journal of Marketing Research 32 (1), 30–41.

Ashford, J. R., Sowden, R. R., 1970. Multi-variate probit analysis. Biometrics 26 (3), 535–546.

Bartlett, M. S., 1953. Approximate confidence intervals. Biometrika 40 (1/2), 12.

Bel, K., Fok, D., Paap, R., 2014. Parameter estimation in multivariate logit models with many binary choices. Econometric Institute Report (Working Paper) 25, 1–24, accepted for Econometric Reviews.

Bel, K., Paap, R., 2013. Modeling the impact of forecast-based regime switching on the inflation rate. Econometric Institute Report (Working Paper) 25, 1–36.

Bel, K., Paap, R., 2014. A multivariate model for multinomial choices. Econometric Institute Report (Working Paper) 26, 1–34.

Bel, K., Paap, R., 2015. Including covariates in two-stage unobserved consideration set discrete choice models, unpublished manuscript, Erasmus University Rotterdam, The Netherlands.

Bel, K., Schoonees, P., 2015. Extending the dale model to multivariate ordered responses using composite likelihood estimation, unpublished manuscript, Erasmus University Rotterdam, The Netherlands.

Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing 12 (1), 9–24.

Ben-Akiva, M., Lerman, S. R., 1985. Discrete choice analysis: Theory and application to travel demand. The MIT Press.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society. Series B (Methodological) 36 (2), 192–236.

Bhat, C. R., Sener, I. N., Eluru, N., 2010. A flexible spatially dependent discrete choice model: Formulation and application to teenagers weekday recreational activity participation. Transportation Research Part B: Methodological 44 (89), 903–921.

Breusch, T. S., 1978. Testing for autocorrelation in dynamic linear models. Australian Economic Papers 17 (31), 334–355.

Bronnenberg, B. J., Vanhonacker, W. R., 1996. Limited choice sets, local price response and implied measures of price competition. Journal of Marketing Research 33 (2), 163–173.

Burda, M., Harding, M., Hausman, J., 2008. A Bayesian mixed logitprobit model for multinomial choice. Journal of Econometrics 147 (2), 232 – 246.

Byrne, J. P., Davis, E. P., 2004. Permanent and temporary inflation uncertainty and investment in the United States. Economics Letters 85 (2), 271–277.

Cameron, A., Trivedi, P., 2005. Microeconometrics methods and applications. Cambridge University Press.

Cappellari, L., Jenkins, S. P., 2006. Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. Stata Journal 6 (2), 156–189.

Carey, V., Zeger, S. L., Diggle, P., 1993. Modelling multivariate binary data with alternating logistic regressions. Biometrika 80 (3), 517–526.

Carro, J. M., 10 2007. Estimating dynamic panel data discrete choice models with fixed effects. Journal of Econometrics 140 (2), 503–528.

Carroll, C. D., 2003. Macroeconomic expectations of households and professional forecasters. The Quarterly Journal of Economics 118 (1), 269–298.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. The Review of Economic Studies 47 (1), 225–238.

Chan, K. S., Tong, H., 1986. On estimating thresholds in autoregressive models. Journal of Time Series Analysis 7 (3), 179–190.

Chib, S., Seetharaman, P., Strijnev, A., 2002. Analysis of multi-category perchase incidence decisions using IRI market basket data. In: Advances in Econometrics. Vol. 16. pp. 75–96.

Christiano, L. J., Gust, C., 2000. The expectations trap hypothesis. Economic Perspectives 24 (2), 21–39.

Clarida, R., Galí, J., Gertler, M., 2000. Monetary policy rules and macroeconomic stability: Evidence and some theory. The Quarterly Journal of Economics 115 (1), 147–180.

Cox, D. R., 1972. The analysis of multivariate binary data. Journal of the Royal Statistical Society. Series C (Applied Statistics) 21 (2), 113–120.

Cox, D. R., Reid, N., 2004. A note on pseudolikelihood constructed from marginal densities. Biometrika 91 (3), 729–737.

Curtin, R. T., 1982. Indicators of consumer behavior: The University of Michigan surveys of consumers. The Public Opinion Quarterly 46 (3), 340–352.

Dale, J. R., 1986. Global cross-ratio models for bivariate, discrete, ordered responses. Biometrics, 909–917.

Davidson, R., MacKinnon, J. G., 2004. Econometric theory and methods. Oxford University Press.

Dixon, W. J., Mood, A. M., 1946. The statistical sign test. Journal of the American Statistical Association 41 (236), 557–566.

Donkers, B., Verhoef, P., Jong, M., 2007. Modeling CLV: A test of competing models in the insurance industry. Quantitative Marketing and Economics 5 (2), 163–190.

Dueker, M. J., Sola, M., Spagnolo, F., 2007. Contemporaneous threshold autoregressive models: Estimation, testing and forecasting. Journal of Econometrics 141 (2), 517–547.

Edwards, Y. D., Allenby, G. M., 2003. Multivariate analysis of multiple response data. Journal of Marketing Research 40 (3), 321–334.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50 (4), 987–987.

Fama, E. F., Gibbons, M. R., 1984. A comparison of inflation forecasts. Journal of Monetary Economics 13 (3), 327 – 348.

Feddag, M.-L., 2013. Composite likelihood estimation for multivariate probit latent traits models. Communications in Statistics - Theory and Methods 42 (14), 2551–2566.

Fellner, W., 1976. Towards a Reconstruction of Macroeceonomics. American Enterprise Institute for Public Policy Research, Washington D.C.

Franses, P. H., Paap, R., 2001. Quantitative Models in Marketing Research. Cambridge University Press.

Gao, X., Song, P., 2010. Composite likelihood bayesian information criteria for model selection in high-dimensional data. Journal of the American Statistical Association 105 (492), 1531–1540.

Gensch, D. H., 1987. A two-stage disaggregate attribute choice model. Marketing Science 6 (3), 223–239.

Geweke, J. F., Keane, M. P., Runkle, D. E., 1997. Statistical inference in the multinomial multiperiod probit model. Journal of Econometrics 80, 125–165.

Givoly, D., Lakonishok, J., 1979. The information content of financial analysts' forecasts of earnings: Some evidence on semi-strong inefficiency. Journal of Accounting and Economics 1 (3), 165–185.

Glonek, G. F. V., McCullagh, P., 1995. Multivariate logistic models. Journal of the Royal Statistical Society. Series B (Methodological) 57 (3), 533–546.

Godambe, V., 1960. An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics 31 (4), 1208–1211.

Godfrey, L. G., 1978. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. Econometrica 46 (6), 1293–1301.

Groen, J. J., Paap, R., Ravazzolo, F., 2013. Real-time inflation forecasting in a changing world. Journal of Business & Economic Statistics 31, 29–44.

Guimares, P., Figueirdo, O., Woodward, D., 2003. A tractable approach to the firm location decision problem. The Review of Economics and Statistics 85 (1), 201–204.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.

Hauser, J. R., Wernerfelt, B., 1990. An evaluation cost model of consideration sets. Journal of Consumer Research 16 (4), 393–408.

Hausman, J. A., Wise, D. A., 1978. A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica (pre-1986) 46 (2), 403.

Hensher, D. A., Greene, W. H., 05 2003. The mixed logit model: The state of practice. Transportation 30 (2), 133.

Honore, B. E., Kyriazidou, E., 2000. Panel data discrete choice models with lagged dependent variables. Econometrica 68 (4), 839–874.

Howard, J. A., Sheth, J. N., 1969. The Theory of Buyter Behavior. New York: John Wiley and Sons.

Joe, H., 2014. Dependence Modeling with Copulas. CRC Press.

Kahn, J. A., McConnell, M. M., Perez-Quiros, G., 2002. On the causes of the increased stability of the U.S. economy. Economic Policy Review - Federal Reserve Bank of New York 8 (1), 183–202.

Kent, J. T., 1982. Robust properties of likelihood ratio test. Biometrika 69 (1), 19–27.

Koop, G., Pesaran, M. H., Potter, S. M., 1996. Impulse response analysis in nonlinear multivariate models. Journal of Econometrics 74 (1), 119–147.

Leduc, S., Sill, K., Stark, T., 2007. Self-fulfilling expectations and the inflation of the 1970s: Evidence from the livingston survey. Journal of Monetary Economics 54 (2), 433–459.

Lee, M.-J., 2002. Panel Data Econometrics: Methods-of-Moments and Limited Dependent Variables. Academic Press, San Diego.

Likert, R., 1932. A technique for the measurement of attitudes. Archives of Psychology 22 (140), 1–55.

Lindsay, B., 1988. Composite likelihood methods. In: Statistical Inference from Stochastic Processes Contemporary Mathematics. Vol. 80. pp. 221–239.

Lucas, Jr, R. E., 1976. Econometric policy evaluation: A critique. Carnegie-Rochester Conference Series on Public Policy 1 (0), 19–46.

Luukkonen, R., Saikkonen, P., Teräsvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrika 75 (3), 491–499.

Maddala, G. S., 1983. Limited-dependent and qualitative variables in econometrics. Vol. 3. Cambridge University Press, Cambridge [etc.].

Manrai, A. K., 1995. Mathematical models of brand choice behavior. European Journal of Operational Research 82 (1), 1 – 17.

McFadden, D., 1978. Modelling the choice of residential location. In: Spatial Interaction Theory and Residential Location. eds. North Holland, Amsterdam, pp. 75–96.

McFadden, D. L., 1983. Handbook of Econometrics. North-Holland Publishing Company, Amsterdam.

Molenberghs, G., Lesaffre, E., 1994. Marginal modeling of correlated ordinal data using a multivariate plackett distribution. Journal of the American Statistical Association 89 (426), 633–644.

Molenberghs, G., Verbeke, G., 2005. Models for discrete longitudinal data. New York NY Springer.

Plackett, R. L., 1965. A class of bivariate distributions. Journal of the American Statistical Association 60 (310), 516–522.

Ramsey, J. B., 1969. Tests for specification errors in classical linear least-squares regression analysis. Journal of the Royal Statistical Society. Series B (Methodological) 31 (2), 350–371.

Roberts, J. H., Lattin, J. M., 1997. Consideration: Review of research and prospects for future insights. Journal of Marketing Research (JMR) 34 (3), 406 – 410.

Rossi, B., Sekhposyan, T., 2010. Have economic models forecasting performance for us output growth and inflation changed over time, and when? International Journal of Forecasting 26 (4), 808–835.

Russell, G. J., Petersen, A., 2000. Analysis of cross category dependence in market basket selection. Journal of Retailing 76 (3), 367–392.

Satterthwaite, F. E., 1946. An approximate distribution of estimates of variance components. Biometrics Bulletin 2 (6), 110–114.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6 (2), 461–464.

Sensier, M., van Dijk, D. J. C., 2004. Testing for volatility changes in US macroeconomic time series. Review of Economics and Statistics 86 (3), 833–839.

Sheridan, J. E., Richards, M. D., Slocum, J. W., 1975. Comparative analysis of expectancy and heuristic models of decision behavior. Journal of Applied Psychology 60 (3), 361–368.

Sinclair, T. M., Gamber, E. N., Stekler, H., Reid, E., 2012. Jointly evaluating the federal reserve forecasts of gdp growth and inflation. International Journal of Forecasting 28 (2), 309–314.

Steiner, C., Groß, A., Entorf, H., 2009. Return and volatility reactions to monthly announcements of business cycle forecasts: An event study based on high-frequency data. ZEW Discussion Papers 09-010, ZEW - Zentrum für Europische Wirtschaftsforschung/Center for European Economic Research.

Stock, J. H., Watson, M. W., 2007. Why has U.S. inflation become harder to forecast? Journal of Money, Credit and Banking 39, 3–33.

Summers, P. M., 2005. What caused the great moderation? Some cross-country evidence. Economic Review - Federal Reserve Bank of Kansas City 90 (3), 5–32,2.

Swait, J. D., 1984. Probabilistic choice set generation in transportation demand models. Ph.D. thesis, Massachusetts Institute of Technology.

Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89 (425), 208–218.

Teräsvirta, T., Anderson, H. M., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics 7, 119–136.

Train, K. E., 2003. Discrete choice methods with simulation. Cambridge University Press, Cambridge [etc.].

van Dijk, D. J. C., 1999. Smooth transition models: Extensions and outlier robust inference. Ph.D. thesis, Erasmus University Rotterdam.

van Dijk, D. J. C., Franses, P. H., 1999. Modeling multiple regimes in the business cycle. Macroeconomic Dynamics 3 (3), 311–340.

van Dijk, D. J. C., Franses, P. H., Boswijk, H. P., 2007. Absorption of shocks in nonlinear autoregressive models. Computational Statistics and Data Analysis 51 (9), 4206–4226.

van Dijk, D. J. C., Teräsvirta, T., Franses, P. H., 2002. Smooth transition autoregressive models: A survey of recent developments. Econometric Reviews 21 (1), 1–47.

van Nierop, E., Bronnenberg, B., Paap, R., Wedel, M., Franses, P. H., 2010. Retrieving unobserved consideration sets from household panel data. Journal of Marketing Research 47 (1), 63 – 74.

Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. Statistica Sinica 21, 5–42.

Vuong, Q. H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57 (2), 307–333.

Wedel, M., 2002. Concomitant variables in finite mixture models. Statistica Neerlandica 56 (3), 362–375.

Wright, P., Barbour, F., 1977. Multiple Criteria Decision Making: Noth-Holland TIMS: Stuides in the Management Science. Amsterdam: North-Holland, Ch. Phased Decisions Strategies: Sequels to Initial Screening, pp. 91–109.

Zhang, X., Boscardin, W. J., Belin, T. R., 2008. Bayesian analysis of multivariate nominal measures using multivariate multinomial probit models. Computational Statistics & Data Analysis 52 (7), 3697–3708.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

579. E. RAVIV *Forecasting Financial and Macroeconomic Variables: Shrinkage , Dimension reduction , and Aggregation*
580. J. TICHEM *Altruism , Conformism , and Incentives in the Workplace*
581. E.S. HENDRIKS *Essays in Law and Economics*
582. X. SHEN *Essays on Empirical Asset Pricing*
583. L.T. GATAREK *Econometric Contributions to Financial Trading , Hedging and Risk Measurement*
584. X. LI *Temporary Price Deviation , Limited Attention and Information Acquisition in the Stock Market*
585. Y. DAI *Efficiency in Corporate Takeovers*
586. S.L. VAN DER STER *Approximate feasibility in real-time scheduling: Speeding up in order to meet deadlines*
587. A. SELIM *An Examination of Uncertainty from a Psychological and Economic Viewpoint*
588. B.Z. YUESHEN *Frictions in Modern Financial Markets and the Implications for Market Quality*
589. D. VAN DOLDER *Game Shows , Gambles , and Economic Behavior*
590. S.P. CEYHAN *Essays on Bayesian Analysis of Time Varying Economic Patterns*
591. S. RENES *Never the Single Measure*
592. D.L. IN 'T VELD *Complex Systems in Financial Economics: Applications to Interbank and Stock Markets*
593. Y. YANG *Laboratory Tests of Theories of Strategic Interaction*

594. M.P. WOJTOWICZ *Pricing Credits Derivatives and Credit Securitization*

595. R.S. SAYAG *Communication and Learning in Decision Making*

596. S.L. BLAUW *Well-to-do or doing well? Empirical studies of wellbeing and development*

597. T.A. MAKAREWICZ *Learning to Forecast: Genetic Algorithms and Experiments*

598. P. ROBALO *Understanding Political Behavior: Essays in Experimental Political Economy*

599. R. ZOUTENBIER *Work Motivation and Incentives in the Public Sector*

600. M.B.W. KOBUS *Economic Studies on Public Facility use*

601. R.J.D. POTTER VAN LOON *Modeling non-standard financial decision making*

602. G. MESTERS *Essays on Nonlinear Panel Time Series Models*

603. S. GUBINS *Information Technologies and Travel*

604. D. KOPÁNYI *Bounded Rationality and Learning in Market Competition*

605. N. MARTYNOVA *Incentives and Regulation in Banking*

606. D. KARSTANJE *Unraveling Dimensions: Commodity Futures Curves and Equity Liquidity*

607. T.C.A.P. GOSENS *The Value of Recreational Areas in Urban Regions*

608. Ł.M. MARĆ *The Impact of Aid on Total Government Expenditures*

609. C. LI *Hitchhiking on the Road of Decision Making under Uncertainty*

610. L. ROSENDAHL HUBER *Entrepreneurship , Teams and Sustainability: a Series of Field Experiments*

611. X. YANG *Essays on High Frequency Financial Econometrics*

612. A.H. VAN DER WEIJDE *The Industrial Organization of Transport Markets: Modeling pricing , Investment and Regulation in Rail and Road Networks*

613. H.E. SILVA MONTALVA *Airport Pricing Policies: Airline Conduct , Price Discrimination , Dynamic Congestion and Network Effects.*

614. C. DIETZ *Hierarchies , Communication and Restricted Cooperation in Cooperative Games*

615. M.A. ZOICAN *Financial System Architecture and Intermediation Quality*

616. G. ZHU *Three Essays in Empirical Corporate Finance*

617. M. PLEUS *Implementations of Tests on the Exogeneity of Selected Variables and their Performance in Practice*

618. B. VAN LEEUWEN *Cooperation , Networks and Emotions: Three Essays in Behavioral Economics*

619. A.G. KOPÁNYI-PEUKER *Endogeneity Matters: Essays on Cooperation and Coordination*

620. X. WANG *Time Varying Risk Premium and Limited Participation in Financial Markets*

621. L.A. GORNICKA *Regulating Financial Markets: Costs and Trade-offs*

622. A. KAMM *Political Actors playing games: Theory and Experiments*

623. S. VAN DEN HAUWE *Topics in Applied Macroeconometrics*

624. F.U. BRÄUNING *Interbank Lending Relationships , Financial Crises and Monetary Policy*

625. J.J. DE VRIES *Estimation of Alonsos Theory of Movements for Commuting*

626. M. POPŁAWSKA *Essays on Insurance and Health Economics*

627. X. CAI *Essays in Labor and Product Market Search*

628. L. ZHAO *Making Real Options Credible: Incomplete Markets , Dynamics , and Model Ambiguity*