Institut für
Volkswirtschaftslehre
und Statistik

No. 596-00

Die Parameterschätzung in fehlspezifizierten Independent Probitmodellen: Eine Monte-Carlo-Studie

Andreas Ziegler

Beiträge zur angewandten Wirtschaftsforschung

Universität Mannheim
A5, 6
D-68131 Mannheim

Die Parameterschätzung in fehlspezifizierten Independent Probitmodellen: Eine Monte-Carlo-Studie

Andreas Ziegler*
Seminar für Statistik

21. Dezember 2000

Zusammenfassung

In dieser Arbeit wird die Parameterschätzung in verschiedenen fehlspezifizierten ein- und mehrperiodigen Mehralternativen-Independent Probitmodellen untersucht. Bei der irrtümlichen Vernachlässigung der zugrunde liegenden kontemporären und intertemporalen Verknüpfungen können sich bei den geschätzten Koeffizienten der erklärenden Variablen systematische Verzerrungen ergeben, falls diese Variablen einen Einfluß auf die Wahlentscheidung besitzen. Bei der inkorrekten Schätzung im einperiodigen Vieralternativen-Independent Probitmodell ist das Ausmaß der Verzerrungen sehr gering. Wesentlich stärkere Auswirkungen zeigen sich bei der Parameterschätzung in fehlspezifizierten Mehrperioden-Mehralternativen-Independent Probitmodellen. Das Ausmaß der Verzerrungen steigt dabei mit zunehmender Anzahl der Perioden und Alternativen.

1 Einleitung

In der Vergangenheit war die Verwendung flexibel formulierter Mehrperioden-Mehralternativen-Probitmodelle (MMPM) bei einer größeren Anzahl an Alternativen und/oder Perioden wegen auftretender Mehrfachintegrale nicht möglich. Erst mit der Entwicklung simulierter Schätzmethoden (vgl. z.B. Lerman/Manski, 1981, McFadden, 1989, Börsch-Supan/Hajivassiliou, 1993, Keane, 1994, Hajivassiliou/McFadden, 1998) ist die Parameterschätzung in derartigen Probitmodellen rechnerisch handhabbar. Allerdings besteht bei der Anwendung von Simulationsschätzverfahren ein hoher Programmieraufwand. Zudem ist die empirische Analyse flexibler multinomialer Probitmodelle häufig durch hohe Rechenzeiten beeinträchtigt.

*Ich danke Prof. Axel Börsch-Supan Ph.D., Elke Eberts, Dr. Angelika Eymann und Prof. Dr. Klaus Winckler für ihre hilfreichen Kommentare.

Auch durch die Nutzung der Programmpakete LIMDEP und GAUSSX, in denen neuerdings die Simulierte Maximum-Likelihood-Methode (SMLM) implementiert ist, kann lediglich der Programmieraufwand vermieden werden.

Darüber hinaus hat sich in vergleichenden Monte-Carlo-Studien gezeigt, daß die Varianz-Kovarianz-Parameter flexibel formulierter mehrperiodiger (binärer oder Mehralternativen-) Probitmodelle mit der SMLM vielfach sehr unpräzise geschätzt werden (vgl. z.B. Keane, 1994, Geweke u.a., 1997, Lee, 1995, 1997a, Hyslop, 1999, Ziegler/Eymann, 2000). Dabei entstehen insbesondere bei der Bestimmung von Autokorrelationskoeffizienten große Verzerrungen. Für eine exaktere und stabilere SMLM-Schätzung derartiger Parameter ist offensichtlich vor allem eine hohe Anzahl an Zufallsziehungen im einbezogenen Simulator notwendig. Mit einer Erhöhung der Anzahl dieser sogenannten Simulationsreplikationen nimmt aber die Rechenzeit der Parameterschätzungen noch weiter zu.

Im Gegensatz zu den Koeffizienten intertemporaler, aber auch kontemporärer Verknüpfungen können die Koeffizienten erklärender Variablen oft schon bei einem geringen Beobachtungsumfang und bei einer geringen Anzahl an Zufallsziehungen stabil und präzise geschätzt werden. Falls nun in empirischen Untersuchungen kein Interesse an der Bestimmung von Varianz-Kovarianz-Parametern besteht, drängt sich entsprechend dieser Ergebnisse bei der Parameterschätzung in diskreten Mehrperioden-Mehralternativen-Entscheidungsmodellen die Vernachlässigung der kontemporären und intertemporalen Korrelationen auf. Mit der Betrachtung einfacher Varianz-Kovarianz-Strukturen wäre letztlich der Einsatz simulierter Schätzmethoden nicht notwendig. Die interessierenden Parameter der erklärenden Variablen könnten stattdessen einfach und schnell mit Hilfe eines konventionellen klassischen Schätzverfahrens, wie z.B. der Maximum-Likelihood-Methode (MLM), bestimmt werden.

Das multinomiale Logitmodell (vgl. McFadden, 1973) hat sich dabei seit einiger Zeit als "Arbeitspferd" der Analyse diskreter Entscheidungen bewährt und wird auch heute noch häufig in empirischen Arbeiten verwendet. Hinsichtlich der Varianz-Kovarianz-Struktur besteht zwischen dem multinomialen Logitmodell und der einfachsten Form des multinomialen Independent Probitmodells eine enge Beziehung. Letzteres zeichnet sich dadurch aus, daß die Varianz-Kovarianz-Matrix der gemeinsam normalverteilten stochastischen Modellkomponenten die Einheitsmatrix darstellt (vgl. auch Hausman/Wise, 1978). Ein solches spezielles Probitmodell beinhaltet also wie das Logitmodell weder kontemporäre noch intertemporale Verknüpfungen. Aufgrund der restriktiven Annahmen zur Korrelationsstruktur taucht in diesen Modellansätzen nicht die Problematik von Vielfachintegralen auf, so daß bei der Parameterschätzung auch bei einer hohen Anzahl an Alternativen und/oder Perioden keine Simulationsmethoden einbezogen werden müssen.

In der vorliegenden Arbeit werden die Auswirkungen der SMLM-Schätzung in Independent Probitmodellen untersucht, falls der datengenerierende Prozeß (DGP) komplexere kontemporäre und intertemporale Verknüpfungen in den stochastischen Modellkomponenten aufweist. Analysiert wird damit die irrtümliche Vernachlässigung der zugrunde liegenden Varianz-Kovarianz-Struktur. Diesen Parameterschätzungen der erklärenden Variablen in fehlspezifizierten multinomialen Probitmodellen werden jeweils die entsprechenden korrekten SMLM-Schätzungen in flexibel formulierten multinomialen Probitmodellen gegenüber gestellt, bei denen die Korrelationsstruktur berücksichtigt werden.

Fehlspezifikationen in Probitmodellen sind bisher in der Literatur nur selten systematisch untersucht worden (vgl. z.B. Lee 1997a, 1997b, Inkmann, 2000, in binären mehrperiodigen Probitmodellen, oder Weeks, 1995, im einperiodigen Dreialternativen-Independent Probitmodell). Insbesondere liegt (nach meiner Kenntnis) noch keine vergleichbare Analyse in mehrperiodigen Mehralternativen-Probitmodellen vor. Daran anknüpfend wird in dieser Arbeit die SMLM-Schätzung in verschiedenen fehlspezifizierten ein- und mehrperiodigen Mehralternativen-Independent Probitmodellen betrachtet. Damit soll insbesondere der Einfluß der Anzahl der Alternativen und Perioden auf die potentiellen Verzerrungen der geschätzten Koeffizienten erklärender Variablen untersucht werden. Darüber hinaus werden die Auswirkungen der Variation des Beobachtungsumfangs auf die Präzision der Parameterschätzungen analysiert.

Konkret ist die vorliegende Arbeit wie folgt strukturiert: Im zweiten Abschnitt wird kurz die SMLM im komplex strukturierten MMPM dargestellt. Im dritten Abschnitt werden die Vorgehensweise der Monte-Carlo-Studien und deren zentralen Ergebnisse beschrieben. Im abschließenden vierten Abschnitt werden im Hinblick auf empirische Arbeiten Schlußfolgerungen aus den Monte-Carlo-Studien gezogen.

2 Simulierte Maximum-Likelihood-Schätzung in Mehrperioden-Mehralternativen-Probitmodellen

Den Monte-Carlo-Studien zur Parameterschätzung im MMPM wird in dieser Arbeit folgende hypothetische Nutzenfunktion einer Beobachtungseinheit i zum Zeitpunkt t bzgl. der Alternative j zugrunde gelegt:

$$v_{ijt} = \gamma' z_{ijt} + \varepsilon_{ijt} \qquad i = 1, \ldots, N; j = 1, \ldots, J; t = 1, \ldots, T$$

Dabei stellen $z_{ijt} = (z_{ijt1}, \ldots, z_{ijtK})'$ einen K-dimensionalen Vektor alternativenspezifischer Attribute und $\gamma = (\gamma_1, \ldots, \gamma_K)'$ den entsprechenden Parametervektor dar. Desweiteren werden die z_{ijt} in den $J \cdot K$-dimensionalen Vektor $z_{it} = (z'_{i1t}, \ldots, z'_{iJt})'$ und dann die z_{it} in den $T \cdot J \cdot K$-dimensionalen Vektor $X_i = (z'_{i1}, \ldots, z'_{iT})'$ einbezogen.

Die stochastische Komponente ε_{ijt} faßt alle nicht beobachteten Faktoren, welche die Entscheidung für eine Alternative j zum Zeitpunkt t beeinflussen, zusammen. Aufgrund der Betrachtung des MMPM gilt:

$$\varepsilon_i = (\varepsilon_{i11}, \ldots, \varepsilon_{iJ1}, \ldots\ldots, \varepsilon_{i1T}, \ldots, \varepsilon_{iJT})' \sim NV(0; \Sigma)$$

Falls Σ durch die Einheitsmatrix gekennzeichnet ist, gelangt man zum Mehrperioden-Mehralternativen-Independent Probitmodell.

Auch der Nutzen υ_{ijt} kann nicht beobachtet werden. Beobachtbar sind dagegen die Realisationen folgender Indikatorvariablen ($i = 1, \ldots, N; j = 1, \ldots, J; t = 1, \ldots, T$):

$$D_{ijt} = \begin{cases} 1 & \text{falls Beobachtungseinheit } i \text{ zum Zeitpunkt } t \text{ Kategorie } j \text{ wählt} \\ 0 & \text{sonst} \end{cases}$$

Im Zeitablauf steht jede Untersuchungseinheit i vor der Wahl zwischen J^T verschiedenen Kategoriensequenzen. Dementsprechend entscheidet sich eine Beobachtungseinheit i im Hinblick auf die tatsächlich gewählte Kategoriensequenz s in jeder Periode t für eine bestimmte Alternative j_{it} (mit $j_{it} = 1, \ldots, J$). Dabei gilt $s \in S$, wobei S die Menge aller J^T potentiellen Kategoriensequenzen darstellt. Somit sind im MMPM die Realisationen folgender Bernoulli-Variablen beobachtbar:

$$Y_{is} = \begin{cases} 1 & \text{falls Beobachtungseinheit } i \text{ die Kategoriensequenz } s \text{ wählt} \\ 0 & \text{sonst} \end{cases}$$

Diese (endogenen) Variablen werden in den J^T-dimensionalen Vektor $Y_i = (Y_{i1}, Y_{i2}, \ldots)'$ einbezogen.

Im folgenden werden alle (freien) Parameter des jeweils betrachteten Probitmodells im Vektor θ zusammengefaßt. Gemäß der stochastischen Nutzenmaximierungshypothese (vgl. z.B. Börsch-Supan, 1987, S. 12 ff, Ronning, 1991, S. 70 ff) entscheidet sich Untersuchungseinheit i zum Zeitpunkt t für Kategorie j_{it}, falls $\upsilon_{ij_{it}t} > \upsilon_{ikt}$ ($i = 1, \ldots, N; k, j_{it} = 1, \ldots, J; k \neq j_{it}; t = 1, \ldots, T$). Daraus lassen sich die Wahrscheinlichkeiten $P_{is}(\theta)$ für die Auswahl einer bestimmten Kategoriensequenz s durch Untersuchungseinheit i ableiten. In Independent Probitmodellen besitzen diese Auswahlwahrscheinlichkeiten eine simple Struktur (vgl. z.B. Hajivassiliou, 2000, S. 93, für den Fall eines einperiodigen multinomialen Probitmodells) und können deshalb problemlos berechnet werden.

Ohne vereinfachende Annahmen bzgl. der Varianz-Kovarianz-Matrix Σ von ε_i sind die Wahrscheinlichkeiten $P_{is}(\theta)$ durch $(J-1) \cdot T$-dimensionale Integrale gekennzeichnet. Im allgemeinen können diese Auswahlwahrscheinlichkeiten schnell und genau mit Hilfe von (stochastischen) Simulationsmethoden, d.h. mit R wiederholten transformierten Ziehungen von Pseudo-Zufallszahlen, approximiert werden. Eine Diskussion verschiedener (erwartungstreuer) Simulationsverfahren findet sich z.B. in Hajivassiliou u.a., 1996, Vijverberg, 1997, oder

Wilde, 1999. In vergleichenden Monte-Carlo-Studien hat sich dabei der sogenannte Geweke-Hajivassiliou-Keane (GHK-)Simulator hinsichtlich der Approximation an die wahre Wahrscheinlichkeit gegenüber anderen Simulationsverfahren als überlegen herauskristallisiert (vgl. auch Börsch-Supan/Hajivassiliou, 1993, Mühleisen, 1994). Mit der Einbeziehung des GHK-Simulators gelangt man zu den simulierten Auswahlwahrscheinlichkeiten $\widetilde{P}_{is}^{GHK}(\theta)$.

Ausgangspunkt der Parameterschätzungen in dieser Arbeit sind N unabhängige Beobachtungspaare (Y_i, X_i). Der wahre, unbekannte und zu schätzende Parametervektor wird dabei speziell mit $\dot{\theta}$ bezeichnet. Somit ergibt sich für den MLM-Schätzer im MMPM:

$$\hat{\theta}_{MLM} = \arg\max_{\theta} \left[\sum_{i=1}^{N} \sum_{s \in S} Y_{is} ln P_{is}(\theta) \right]$$

In Mehrperioden-Mehralternativen-Independent Probitmodellen ist die MLM-Schätzung auch für große J und T rechnerisch handhabbar. Im flexibel formulierten MMPM ist dagegen für jede Beobachtungseinheit i und für jede Iteration im Maximierungsprozeß die Berechnung eines $(J-1) \cdot T$-dimensionalen Integrals zur Bestimmung der Auswahlwahrscheinlichkeiten $P_{is}(\theta)$ erforderlich. Diese Wahrscheinlichkeiten können mit einem Simulationsverfahren approximiert werden. Durch die Verknüpfung der MLM mit einem Simulator gelangt man zur SMLM (vgl. z.B. Gouriéroux/Monfort, 1996, S. 41 ff). Mit der Einbeziehung des GHK-Simulators $\widetilde{P}_{is}^{GHK}(\theta)$ in obigen MLM-Ansatz ergibt sich im MMPM folgender SMLM/GHK-Schätzer:

$$\hat{\theta}_{SMLM}^{GHK} = \arg\max_{\theta} \left[\sum_{i=1}^{N} \sum_{s \in S} Y_{is} ln \widetilde{P}_{is}^{GHK}(\theta) \right]$$

Zu beachten ist, daß in den Monte-Carlo-Studien im nächsten Abschnitt sowohl im Rahmen der flexibel formulierten Probitmodelle als auch im Rahmen der Independent Probitmodelle SMLM/GHK-Schätzungen vorgenommen werden. Damit können die allein aus der Modellfehlspezifikation resultierenden Effekte auf die Schätzergebnisse untersucht werden.

3 Monte-Carlo-Studie

3.1 Design

Im folgenden werden Parameterschätzungen in verschieden strukturierten fehlspezifizierten ein- und mehrperiodigen Mehralternativen-Independent Probitmodellen miteinander verglichen. Exemplarisch werden einperiodige Vieralternativen-, fünfperiodige Dreialternativen- sowie achtperiodige Vieralternativen-Probitmodelle betrachtet. Den SMLM/GHK-Schätzungen in fehlspezifizierten Probitmodellen werden jeweils die entsprechenden SMLM/GHK-Schätzungen der Parameter der erklärenden Variablen in korrekt spezifizierten Probitmodellen gegenüber gestellt (letztere sind ausführlich in Ziegler/Eymann, 2000, dargestellt).

Ausgangspunkt der Betrachtungen ist ein einperiodiges Vieralternativen-Probitmodell. Die Anzahl der Replikationen des DGP beträgt hierbei 200. Diesem einperiodigen Mehralternativen-Probitmodell werden zwei mehrperiodige Mehralternativen-Probitmodelle gegenüber gestellt. Die Auswahl eines fünfperiodigen Dreialternativen-Probitmodells erfolgt in erster Linie aufgrund der noch akzeptablen Rechenzeiten bei der SMLM/GHK-Schätzung im flexibel formulierten MMPM über die 200 Replikationen des DGP. Mit der Betrachtung eines achtperiodigen Vieralternativen-Probitmodells können einerseits (gegenüber dem einperiodigen Vieralternativen-Probitmodell) die Auswirkungen einer ausschließlichen Erhöhung von T und andererseits (gegenüber dem fünfperiodigen Dreialternativen-Probitmodell) die Effekte einer simultanen Erhöhung von T und J analysiert werden. Allerdings wird der DGP in diesem Fall wegen der hohen Rechenzeiten bei der SMLM/GHK-Schätzung im flexibel formulierten MMPM lediglich 20 mal repliziert.

In allen untersuchten ein- und mehrperiodigen Mehralternativen-Probitmodellen werden zwei alternativenspezifische erklärende Variablen einbezogen (d.h. $K = 2$). Damit wird folgendes einperiodige Vieralternativen-Probitmodell betrachtet ($i = 1, \ldots, N$):

$$
\begin{aligned}
v_{i11} &= \gamma_1 z_{i111} + \gamma_2 z_{i112} + \varepsilon_{i11} \\
v_{i21} &= \gamma_1 z_{i211} + \gamma_2 z_{i212} + \varepsilon_{i21} \\
v_{i31} &= \gamma_1 z_{i311} + \gamma_2 z_{i312} + \varepsilon_{i31} \\
v_{i41} &= \gamma_1 z_{i411} + \gamma_2 z_{i412} + \varepsilon_{i41}
\end{aligned}
$$

Für die zwei alternativenspezifischen Attribute gilt dabei ($i = 1, \ldots, N; j = 1, \ldots, 4$):

$$z_{ij11} \sim NV(0;2) \quad z_{ij21} \sim NV(0;2)$$

Den entsprechenden Parametern liegen folgende Werte zugrunde:

$$\dot{\gamma}_1 = 1 \quad \dot{\gamma}_2 = 0$$

Im fünfperiodigen Dreialternativen-Probitmodell gilt ($i = 1, \ldots, N;\ t = 1, \ldots, 5$):

$$
\begin{aligned}
v_{i1t} &= \gamma_1 z_{i1t1} + \gamma_2 z_{i1t2} + \varepsilon_{i1t} \\
v_{i2t} &= \gamma_1 z_{i2t1} + \gamma_2 z_{i2t2} + \varepsilon_{i2t} \\
v_{i3t} &= \gamma_1 z_{i3t1} + \gamma_2 z_{i3t2} + \varepsilon_{i3t}
\end{aligned}
$$

Dabei werden im Hinblick auf intertemporale Verknüpfungen folgende zwei alternativenspezifische Attribute betrachtet ($i = 1, \ldots, N; j = 1, \ldots, 3; t = 1, \ldots, 5$):

$$z_{ijt1} = z_{ij1}^{(1)} + z_{ijt1}^{(2)} \quad \text{wobei} \quad z_{ij1}^{(1)} \sim NV(0;1) \quad \text{und} \quad z_{ijt1}^{(2)} \sim NV(0;1)$$

$$z_{ijt2} = z_{ij2}^{(1)} + z_{ijt2}^{(2)} \quad \text{wobei} \quad z_{ij2}^{(1)} \sim NV(0;1) \quad \text{und} \quad z_{ijt2}^{(2)} \sim NV(0;1)$$

Den entsprechenden Parametern liegen folgende Werte zugrunde:

$\dot{\gamma}_1 = 1 \quad \dot{\gamma}_2 = 0$

Im achtperiodigen Vieralternativen-Probitmodell gilt ($i = 1, \ldots, N$; $t = 1, \ldots, 8$):

$$\begin{aligned}
\upsilon_{i1t} &= \gamma_1 z_{i1t1} + \gamma_2 z_{i1t2} + \varepsilon_{i1t} \\
\upsilon_{i2t} &= \gamma_1 z_{i2t1} + \gamma_2 z_{i2t2} + \varepsilon_{i2t} \\
\upsilon_{i3t} &= \gamma_1 z_{i3t1} + \gamma_2 z_{i3t2} + \varepsilon_{i3t} \\
\upsilon_{i4t} &= \gamma_1 z_{i4t1} + \gamma_2 z_{i4t2} + \varepsilon_{i4t}
\end{aligned}$$

Hinsichtlich der beiden alternativenspezifischen Attribute wird hier auch eine Dummy-Variable in die Betrachtung einbezogen, d.h. ($i = 1, \ldots, N; j = 1, \ldots, 4; t = 1, \ldots, 8$):

$$z_{ijt1} = z_{ij1}^{(1)} + z_{ijt1}^{(2)} \quad \text{wobei} \quad z_{ij1}^{(1)} \sim NV(0;1) \quad \text{und} \quad z_{ijt1}^{(2)} \sim NV(0;1)$$

$$z_{ijt2} \sim BV(0.5)$$

Den entsprechenden Parametern liegen folgende Werte zugrunde:

$\dot{\gamma}_1 = 1 \quad \dot{\gamma}_2 = 1$

Der DGP enthält dabei im einperiodigen Vieralternativen-Probitmodell kontemporäre Verknüpfungen und in den beiden mehrperiodigen Mehralternativen-Probitmodellen kontemporäre, zeitinvariante und autoregressive Verknüpfungen in den stochastischen Modellkomponenten. Die konkrete Korrelationsstruktur sowie die vorgenommenen Restriktionen zur formalen Identifikation der Modelle sind ausführlich in Ziegler/Eymann, 2000, beschrieben.

Im einperiodigen Vieralternativen-Probitmodell wird der Beobachtungsumfang zwischen $N = 1000$ bzw. $N = 2000$, in den beiden mehrperiodigen Mehralternativen-Probitmodellen wird der Beobachtungsumfang zwischen $N = 250$ bzw. $N = 500$ variiert. Die ausgewiesenen Resultate basieren auf einer Anzahl von $R = 10$ Simulationsreplikationen im GHK-Simulator.

Im folgenden gehen in die Statistik *Bias* die durchschnittlichen Verzerrungen und in die Statistik *Rmse* die Wurzel aus der mittleren quadratischen Abweichung der SMLM/GHK-Schätzwerte über die einzelnen Replikationen des DGP jeweils bezogen auf die wahren Parameter der erklärenden Variablen ein. Damit beinhaltet *Rmse* auch die Standardabweichungen (*Stab*) der Schätzwerte über die 20 bzw. 200 Replikationen des DGP. Neben *Bias*, *Stab* und *Rmse* werden das arithmetische Mittel ($\bar{\hat{\theta}}$), das Minimum (Min), der Median (Med), sowie das Maximum (Max) der SMLM/GHK-Schätzwerte über die einzelnen Replikationen des DGP ausgewiesen. Im einperiodigen Vieralternativen- und im fünfperiodigen Dreialternativen-Probitmodell werden darüber hinaus die 25%- bzw. 75%-Quantile ($25\%, 75\%$) über alle 200 Replikationen des DGP betrachtet.

3.2 Ergebnisse

3.2.1 Einperiodiges Vieralternativen-Probitmodell

In Tabelle 1 sind die ausführlichen Statistiken zur SMLM/GHK-Schätzung der Koeffizienten der erklärenden Variablen im einperiodigen Vieralternativen-Probitmodell abgebildet. Im oberen Teil der Tabelle wird mit der korrekten Schätzung im flexibel formulierten Probitmodell die zugrunde liegende Varianz-Kovarianz-Struktur berücksichtigt. Dabei ergeben sich stabile Parameterschätzungen mit geringen Verzerrungen. Mit der Zunahme des Beobachtungsumfangs von $N = 1000$ auf $N = 2000$ vermindert sich aufgrund der sinkenden Streuungen sowohl bei γ_1 als auch bei γ_2 der Wert von $Rmse$. Wegen der zugrunde gelegten Parameterwerte (im DGP gilt $\dot{\gamma}_1 = 1$ sowie $\dot{\gamma}_2 = 0$) und wegen der jeweils vorgegebenen Parameterstartwerte von Null für den iterativen Maximierungsprozeß der SMLM/GHK wird γ_2 vergleichsweise mit geringeren Werten von $Stab$ und damit $Rmse$ geschätzt.

Im unteren Teil von Tabelle 1 werden irrtümlich die kontemporären Verknüpfungen nicht bei der SMLM/GHK-Schätzung einbezogen. Dabei zeigt sich, daß der Parameter γ_1 unabhängig von N im Durchschnitt überschätzt wird. Allerdings sind die resultierenden Verzerrungen sehr moderat. Aufgrund der geringen Streuungen ergeben sich für γ_1 und mit $N = 1000$ bei der SMLM/GHK-Schätzung im fehlspezifizierten einperiodigen Vieralternativen-Independent Probitmodell sogar geringere Werte von $Rmse$ als bei der korrekten SMLM/GHK-Schätzung im flexibel formulierten einperiodigen Vieralternativen-Probitmodell. Darüber hinaus sind beim Parameter γ_2 überhaupt keine systematischen Verzerrungen zu erkennen. Anknüpfend an die Bemerkungen in der Einleitung dieser Arbeit sprechen auch diese Ergebnisse dafür, daß die Einbeziehung der Schätzung von Varianz-Kovarianz-Parametern für die Bestimmung der Koeffizienten der erklärenden Variablen nur von untergeordneter Bedeutung ist.

3.2.2 Fünfperiodiges Dreialternativen-Probitmodell

Tabelle 2 enthält die ausführlichen Statistiken zur SMLM/GHK-Schätzung der Parameter der erklärenden Variablen im fünfperiodigen Dreialternativen-Probitmodell. Im oberen Teil der Tabelle werden bei der korrekten Schätzung im flexibel formulierten MMPM die kontemporären, zeitinvarianten und autoregressiven Korrelationen berücksichtigt. Wie bei der Schätzung im flexibel formulierten einperiodigen Vieralternativen-Probitmodell werden auch hier beide Koeffizienten γ_1 und γ_2 stabil und im Durchschnitt unverzerrt geschätzt. Erneut sinken mit zunehmendem Beobachtungsumfang N die Werte von $Stab$ und somit von $Rmse$. Darüber hinaus ist wiederum die Güte der SMLM/GHK-Schätzung von γ_2 vergleichsweise höher.

Im unteren Teil von Tabelle 2 wird die zugrunde gelegte Varianz-Kovarianz-Struktur irrtüm-

Tabelle 1: SMLM/GHK-Schätzung der Parameter der erklärenden Variablen im einperiodigen Vieralternativen-Probitmodell

Schätzung: Einbeziehung kontemporärer Verknüpfungen										
	θ	$\bar{\hat{\theta}}$	*Bias*	*Min*	25%	*Med*	75%	*Max*	*Stab*	*Rmse*
$N = 1000$	γ_1	1.0208	0.0208	0.8176	0.9579	1.0129	1.0782	1.2778	0.0888	0.0912
	γ_2	0.0006	0.0006	-0.0537	-0.0104	-0.0009	0.0117	0.0471	0.0169	0.0169
$N = 2000$	γ_1	1.0066	0.0066	0.8917	0.9617	0.9994	1.0343	1.1836	0.0560	0.0564
	γ_2	0.0000	0.0000	-0.0420	-0.0078	0.0001	0.0076	0.0290	0.0112	0.0112
Schätzung: Independent Probitmodell										
	θ	$\bar{\hat{\theta}}$	*Bias*	*Min*	25%	*Med*	75%	*Max*	*Stab*	*Rmse*
$N = 1000$	γ_1	1.0604	0.0604	0.9482	1.0270	1.0545	1.0884	1.2195	0.0510	0.0791
	γ_2	-0.0004	-0.0004	-0.0626	-0.0133	0.0005	0.0121	0.0512	0.0180	0.0180
$N = 2000$	γ_1	1.0502	0.0502	0.9683	1.0295	1.0493	1.0711	1.1406	0.0317	0.0595
	γ_2	-0.0000	-0.0000	-0.0417	-0.0086	0.0004	0.0088	0.0287	0.0119	0.0119

Tabelle 2: SMLM/GHK-Schätzung der Parameter der erklärenden Variablen im fünfperiodigen Dreialternativen-Probitmodell

Schätzung: Einbeziehung kontemporärer und intertemporaler Verknüpfungen										
	θ	$\bar{\hat{\theta}}$	*Bias*	*Min*	25%	*Med*	75%	*Max*	*Stab*	*Rmse*
$N = 250$	γ_1	1.0037	0.0037	0.7947	0.9387	1.0040	1.0627	1.1857	0.0835	0.0835
	γ_2	-0.0053	-0.0053	-0.0870	-0.0256	-0.0047	0.0158	0.0752	0.0323	0.0327
$N = 500$	γ_1	0.9972	-0.0028	0.8599	0.9468	0.9952	1.0338	1.1811	0.0608	0.0609
	γ_2	-0.0021	-0.0021	-0.0644	-0.0155	-0.0018	0.0120	0.0600	0.0221	0.0222
Schätzung: Independent Probitmodell										
	θ	$\bar{\hat{\theta}}$	*Bias*	*Min*	25%	*Med*	75%	*Max*	*Stab*	*Rmse*
$N = 250$	γ_1	0.7120	-0.2880	0.6059	0.6793	0.7089	0.7435	0.8302	0.0415	0.2917
	γ_2	-0.0071	-0.0071	-0.0758	-0.0250	-0.0047	0.0124	0.0615	0.0282	0.0291
$N = 500$	γ_1	0.7127	-0.2873	0.6453	0.6911	0.7115	0.7291	0.7856	0.0285	0.2894
	γ_2	-0.0027	-0.0027	-0.0564	-0.0163	-0.0023	0.0114	0.0586	0.0206	0.0207

lich vernachlässigt. Bei der SMLM/GHK-Schätzung von γ_2 im fehlspezifizierten fünfperiodigen Dreialternativen-Independent Probitmodell sind im Durchschnitt erneut keine Verzerrungen zu erkennen. Dieses Ergebnis deckt sich mit demjenigen bei der inkorrekten Parameterschätzung im einperiodigen Vieralternativen-Independent Probitmodell. Offensichtlich ist bei der SMLM/GHK-Schätzung des Koeffizienten einer alternativenspezifischen erklärenden Variablen im fehlspezifizierten Independent Probitmodell lediglich mit geringen bzw. mit überhaupt keinen systematischen Verzerrungen zu rechnen, falls diese erklärende Variable tatsächlich keinen Einfluß auf die Wahlentscheidung ausübt (d.h. falls der entsprechende Parameter im DGP den Wert Null besitzt).

Das wesentliche Resultat im unteren Teil von Tabelle 2 ist die starke und ausnahmslose Unterschätzung von γ_1 über alle 200 Replikationen des durch kontemporäre, zeitinvariante und autoregressive Verknüpfungen gekennzeichneten DGP. Dieses Ergebnis zeigt sich unabhängig vom Beobachtungsumfang N. Bemerkenswert ist, daß demgegenüber γ_1 bei der inkorrekten Parameterschätzung im einperiodigen Vieralternativen-Independent Probitmodell durchschnittlich überschätzt wird. Vor allem aber sind die absoluten Verzerrungen bei der SMLM/GHK-Schätzung von γ_1 im betrachteten fehlspezifizierten mehrperiodigen Mehralternativen-Independent Probitmodell im Vergleich von deutlich höherem Ausmaß. Damit sollte (entgegen der Argumentation in Abschnitt 3.2.1) zur Vermeidung systematischer Verzerrungen auch bei der alleinigen Schätzung der Koeffizienten erklärender Variablen die korrekte Varianz-Kovarianz-Struktur einbezogen werden. Dadurch wird aber wegen des Auftauchens von Mehrfachintegralen in der Regel der Einsatz von Simulationsschätzverfahren wie der SMLM/GHK notwendig.

3.2.3 Achtperiodiges Vieralternativen-Probitmodell

Die letzte Überlegung wird durch die Betrachtung des achtperiodigen Vieralternativen-Probitmodells bekräftigt. In Tabelle 3 sind die ausführlichen Statistiken zur SMLM/GHK-Schätzung der Parameter der erklärenden Variablen in diesem MMPM abgebildet. Im oberen Teil erfolgt die Einbeziehung der kontemporären, zeitinvarianten und autoregressiven Verknüpfungen. Auch bei dieser korrekten SMLM/GHK-Schätzung im flexibel formulierten achtperiodigen Vieralternativen-Probitmodell ergeben sich stabile Parameterschätzungen mit sehr geringen Verzerrungen. Erneut vermindert sich bei wachsendem N wegen des sinkenden *Stab* auch *Rmse*, insbesondere beim Koeffizienten γ_2. Festzuhalten ist dabei, daß die Schätzungen des Parameters γ_1 der betrachteten normalverteilten erklärenden Variablen im Vergleich zu den Schätzungen des Parameters γ_2 der betrachteten Dummy-Variablen geringere Streuungen aufweisen.

Zu beachten ist im Unterschied zur Analyse im einperiodigen Vieralternativen- bzw. im fünfperiodigen Dreialternativen-Probitmodell, daß im DGP sowohl $\dot{\gamma}_1 = 1$ als auch $\dot{\gamma}_2 = 1$ gelten. Dadurch ergeben sich aber bei der Schätzung beider Koeffizienten im fehlspezifizierten achtperiodigen Vieralternativen-Independent Probitmodell systematische Verzerrungen.

Tabelle 3: SMLM/GHK-Schätzung der Parameter der erklärenden Variablen im achtperiodigen Vieralternativen-Probitmodell

Schätzung: Einbeziehung kontemporärer und intertemporaler Verknüpfungen								
	θ	$\bar{\hat{\theta}}$	$Bias$	Min	Med	Max	$Stab$	$Rmse$
$N = 250$	γ_1	1.0243	0.0243	0.9249	1.0076	1.1882	0.0672	0.0717
	γ_2	1.0222	0.0222	0.8010	1.0062	1.3166	0.1197	0.1219
$N = 500$	γ_1	0.9973	-0.0027	0.9435	0.9795	1.0977	0.0460	0.0460
	γ_2	0.9899	-0.0101	0.8691	0.9823	1.1543	0.0706	0.0713
Schätzung: Independent Probitmodell								
	θ	$\bar{\hat{\theta}}$	$Bias$	Min	Med	Max	$Stab$	$Rmse$
$N = 250$	γ_1	0.6610	-0.3390	0.5835	0.6598	0.7053	0.0336	0.3494
	γ_2	0.6430	-0.3570	0.5288	0.6429	0.7403	0.0547	0.3703
$N = 500$	γ_1	0.6436	-0.3564	0.6154	0.6385	0.6947	0.0213	0.3663
	γ_2	0.6339	-0.3661	0.5562	0.6339	0.6882	0.0336	0.3771

Die entsprechenden Ergebnisse sind im unteren Teil von Tabelle 3 ausgewiesen. Das wesentliche Resultat ist die unabhängig von N sehr starke und ausnahmslose Unterschätzung von γ_1 und von γ_2 über alle 20 Replikationen des durch kontemporäre, zeitinvariante und autoregressive Korrelationen gekennzeichneten DGP. Zu betonen ist, daß das Ausmaß der durchschnittlichen Verzerrungen bei beiden Parametern γ_1 und γ_2 jeweils höher ist als das entsprechende Ausmaß bei der inkorrekten Schätzung von γ_1 im fünfperiodigen Dreialternativen-Independent Probitmodell. Insofern ist mit der Zunahme der Anzahl J der Alternativen sowie der Anzahl T der Perioden eine Verschärfung der Effekte einer Parameterschätzung im fehlspezifizierten Mehrperioden-Mehralternativen-Independent Probitmodell festzustellen.

4 Schlußfolgerungen

Die Monte-Carlo-Studien in dieser Arbeit offenbaren zum Teil massive Auswirkungen der SMLM/GHK-Schätzung in fehlspezifizierten Mehralternativen-Independent Probitmodellen. Die bei den Koeffizienten der (alternativenspezifischen) erklärenden Variablen (deren Wert im DGP von Null abweicht) auftauchenden Verzerrungen werden dabei nicht vom Beobachtungsumfang N beeinflußt. Im Umkehrschluß bedeutet dies für empirische Arbeiten, daß derartige Verzerrungen mit einer Erhöhung von N nicht eingedämmt werden können, falls bei der Parameterschätzung irrtümlich kontemporäre sowie (im mehrperiodigen Fall) sämtliche intertemporale Verknüpfungen unberücksichtigt bleiben. Dieser Sachverhalt ergibt sich bei der SMLM/GHK-Schätzung aller betrachteten fehlspezifizierten ein- und mehrperiodigen Mehralternativen-Independent Probitmodelle und damit unabhängig von der Anzahl J der Alternativen sowie von der Anzahl T der Perioden.

Hinsichtlich des Ausmaßes der Verzerrungen spielen dagegen J und T eine wichtige Rolle. So liegen bei der Schätzung der Koeffizienten (alternativenspezifischer) erklärender Variablen im fehlspezifizierten einperiodigen Vieralternativen-Independent Probitmodell vergleichsweise moderate durchschnittliche Verzerrungen vor. Dagegen ist die Stärke der systematischen Verzerrungen bei der inkorrekten SMLM/GHK-Schätzung in mehrperiodigen Mehralternativen-Independent Probitmodellen deutlich höher. Die stärksten Auswirkungen erscheinen demnach bei der Parameterschätzung im fehlspezifizierten achtperiodigen Vieralternativen-Independent Probitmodell, d.h. also bei hohen J und T.

Zu betonen ist dabei, daß sich diese SMLM/GHK-Schätzergebnisse lediglich auf diejenigen (alternativenspezifischen) erklärenden Variablen beziehen, die auch tatsächlich einen Einfluß auf die Wahlentscheidung besitzen. Falls dagegen ein derartiger Koeffizient im DGP den Wert Null aufweist, ist unabhängig von J, T, N und R bei der fehlspezifizierten Schätzung dieser Koeffizienten nicht mit systematischen Verzerrungen zu rechnen.

Allerdings gilt dieses Resultat offensichtlich nur für die Parameter alternativenspezifischer erklärender Variablen. Weiter führende eigene Versuche haben gezeigt, daß bei der SMLM/GHK-Schätzung von Koeffizienten individuenspezifischer erklärender Variablen, die im DGP den Wert Null besitzen, sehr wohl Verzerrungen auftreten können, falls eine inkorrekte Varianz-Kovarianz-Struktur zugrunde gelegt wird. Eine systematische Analyse hinsichtlich der fehlspezifizierten Schätzung auch dieser Parametergruppe wäre für die Zukunft wünschenswert, zumal in empirischen Arbeiten häufig individuenspezifische erklärende Variablen einbezogen werden.

Aus den vorliegenden Untersuchungen zeigt sich letztlich, daß auch bei der allein interessierenden Schätzung der Parameter der erklärenden Variablen die zugrunde liegende Varianz-Kovarianz-Struktur korrekt erfaßt werden sollte (vgl. dazu auch Kaltenborn, 1997, S. 114 ff). Damit ergibt sich aber beim Auftreten von Mehrfachintegralen die Notwendigkeit der Einbeziehung von Simulationsmethoden. Lediglich im Rahmen eines einperiodigen Probitmodells mit einer kleinen Anzahl J an Alternativen scheint die Berücksichtigung der kontemporären Verknüpfungen für die präzise Schätzung der Koeffizienten alternativenspezifischer erklärender Variablen nicht sehr entscheidend zu sein. Allerdings ist zu berücksichtigen, daß z.B. für $J = 3$ oder $J = 4$ auch bei der Parameterschätzung im flexibel formulierten (einperiodigen) Probitmodell die Anwendung der konventionellen MLM möglich wäre. Damit ergibt sich in diesen Fällen auch nicht das Problem eines hohen Programmier- und Rechenzeitaufwandes.

Hohe Rechenzeiten entstehen bei der simulierten Parameterschätzung im flexibel formulierten MMPM vor allem bei hohen J und/oder T. Dennoch sollte gerade in diesen Fällen für die präzise Schätzung der Parameter der erklärenden Variablen die korrekte Varianz-Kovarianz-Struktur berücksichtigt werden, da sich bei hohen J und T mit der Schätzung im fehlspezifizierten Independent Probitmodell die stärksten Verzerrungen ergeben. Zu betonen ist, daß die SMLM/GHK seit kurzem in den Programmpaketen GAUSSX und LIMDEP implementiert ist, so daß die Eintrittsbarrieren für die empirische Analyse flexibler multinomialer Probitmodelle stark reduziert sind. Dabei sollte sich der potentielle Anwender auch nicht von den (bei großen J und T) anfallenden hohen Rechenzeiten abschrecken lassen, zumal nur mit der Schätzung im flexibel formulierten MMPM die für viele empirische Arbeiten interessanten Varianz-Kovarianz-Parameter bestimmt werden können.

Literatur

Börsch-Supan, A. (1987). *Econometric Analysis of Discrete Choice. With Applications on the Demand for Housing in the U.S. and West-Germany.* Springer-Verlag, Berlin u.a.

Börsch-Supan, A., Hajivassiliou, V. A. (1993). Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models. *Journal of Econometrics* **58** 347-368.

Geweke, J., Keane, M., Runkle, D. (1997). Statistical Inference in the Multinomial Multiperiod Probit Model. *Journal of Econometrics* **80** 125-165.

Gouriéroux, C., Monfort, A. (1996). *Simulation-Based Econometric Methods.* Oxford University Press.

Hajivassiliou, V. A. (2000). Some Practical Issues in Maximum Simulated Likelihood. In: Mariano, R., Schuermann, T., Weeks, M. J. (Hrsg.). *Simulation-Based Inference in Econometrics: Methods and Applications.* Cambridge University Press, Cambridge.

Hajivassiliou, V. A., McFadden, D. (1998). The Method of Simulated Scores for the Estimation of LDV Models. *Econometrica* **66** (4) 863-896.

Hajivassiliou, V. A., McFadden, D., Ruud, P. (1996). Simulation of Multivariate Normal Rectangle Probabilities and their Derivations. Theoretical and Computational Results. *Journal of Econometrics* **72** 85-134.

Hausman, J. A., Wise, D. A. (1978). A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. *Econometrica* **46** (2) 403-426.

Hyslop, D. R. (1999). State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women. *Econometrica* **67** (6) 1255-1294.

Inkmann J. (2000). Misspecified Heteroscedasticity in the Panel Probit Model: A Small Sample Comparison of GMM and SML Estimators. *Journal of Econometrics* **97** 227-259.

Kaltenborn, U. (1997). Die Anwendung simulativer Schätzverfahren für Discrete-Choice-Modelle am Beispiel von Daten des Sozioökonomischen Panels. *Dissertation.* Fachbereich Wirtschaftswissenschaft der Freien Universität Berlin.

Keane, M. (1994). A Computationally Practical Simulation Estimator for Panel Data. *Econometrica* **62** (1) 95-116.

Lee, L.-F. (1995). Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models. *Econometric Theory* **11** 437-483.

Lee, L.-F. (1997a). Simulated Maximum Likelihood Estimation of Dynamic Discrete Choice Statistical Models. Some Monte Carlo Results. *Journal of Econometrics* **82** 1-35.

Lee L.-F. (1997b). Some Common Structures of Simulated Specification Tests in Multinormal Discrete and Limited Dependent Variables Models. *Working Paper No. 97-4, The Hong Kong University of Science and Technology, Department of Economics.*

Lerman, S. R., Manski, C. F. (1981). On the Use of Simulated Frequencies to Approximate Choice Probabilities. In: Manski, C. F., McFadden, D. (Hrsg.). *Structural Analysis of Discrete Data with Econometric Applications.* 305-319, MIT Press, Cambridge/London.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka P. (Hrsg.). *Frontiers in Econometrics.* 105-142, Academic Press, New York.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration. *Econometrica* **57** (5) 995-1026.

Mühleisen, M. (1994). *Human Capital Decay and Persistence. A Simulation Approach to German Unemployment.* Campus Verlag, Frankfurt am Main.

Ronning, G. (1991). *Mikroökonometrie.* Springer-Verlag, Berlin u.a.

Vijverberg, W. P. M. (1997). Monte Carlo Evaluation of Multivariate Normal Probabilities. *Journal of Econometrics* **76** 281-307.

WEEKS, M. (1995). Circumventing the Curse of Dimensionality in Applied Work Using Computer Intensive Methods. *The Economic Journal* **105** 520-530.

WILDE, J. (1999). *Gemischte simultane Modelle für Querschnittsdaten.* Lang, Frankfurt am Main.

ZIEGLER, A., EYMANN, A. (2000). Zur Simulated Maximum-Likelihood-Schätzung von Mehrperioden-Mehralternativen-Probitmodellen. Erscheint in: *Allgemeines Statistisches Archiv.*