

UNIVERSITÄT POTSDAM  
Wirtschafts- und Sozialwissenschaftliche Fakultät

**STATISTISCHE DISKUSSIONSBEITRÄGE**

Nr. 2

Wolfram Kempe

**Das Arbeitsangebot verheirateter Frauen  
in den neuen und alten Bundesländern**

- Eine semiparametrische Regressionsanalyse -



Potsdam 1996  
ISSN 0949-068X

# STATISTISCHE DISKUSSIONSBEITRÄGE

Herausgeber: Hans Gerhard Strohe

Nr. 2

Wolfram Kempe

## Das Arbeitsangebot verheirateter Frauen in den neuen und alten Bundesländern

- Eine semiparametrische Regressionsanalyse -

Lehrstuhl Statistik  
der Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der Universität Potsdam  
Postfach 90 03 27  
D-14439 Potsdam  
Tel. (+49 331) 977-32 25  
Fax. (+49 331) 977-32 10  
e-mail: strohe@rz.uni-potsdam.de

1996  
ISSN 0949-068X

Wolfram Kempe

# Das Arbeitsangebot verheirateter Frauen in den neuen und alten Bundesländern – Eine semiparametrische Regressionsanalyse –

## Zusammenfassung

In diesem Beitrag wird eine Regressionsanalyse vorgestellt, die die Einflüsse auf die Entscheidung verheirateter deutscher Frauen untersucht, eine Erwerbstätigkeit aufzunehmen. Um Differenzen im Verhalten von ost- und westdeutschen Frauen zu ermitteln, erfolgte die Untersuchung getrennt in zwei Datensätzen. Zur Vermeidung von Annahmen über die Art des Zusammenhanges wurde das Generalisierte Additive Modell (GAM) gewählt, ein semiparametrisches Regressionsmodell. Diese Modellform, die nichtparametrische und parametrische Regressionsmethoden in sich vereint, hat bisher wenig Verbreitung in der Praxis gefunden. Dies lag vor allem am Schätzverfahren, dem Backfitting. Seit etwa einem Jahr gibt es neue Ansätze, in dieser Modellform zu schätzen. Die analytischen Eigenschaften des neuen Schätzers lassen sich leichter bestimmen. Mit dieser Schätzung konnten Unterschiede zwischen Ost und West genau herausgearbeitet werden und die funktionalen Zusammenhänge zwischen Einflußvariablen und Antwortvariable untersucht werden. Die Analyse brachte deutliche Unterschiede im Erwerbsverhalten zwischen der Frauen beider Landesteile zum Vorschein.

Schlüsselwörter: Arbeitsangebot, Frauenerwerbstätigkeit, nichtparametrische Regression, GAM, Integrationsschätzer

## Abstract

This paper will focus on the regression analysis of labor supply decisions of married German women. In order to determine differences East and West German women were compared separately. To avoid assumptions about the functional type of correlation the Generalized Additive Model, a semiparametric regression model, was chosen. So far, this pattern consisting of nonparametric and parametric methods has not found acceptance in practical application. Reason for that is the backfitting-estimator. One year ago new ideas for the estimation by GAM were found. The analytical features of the new estimator are easier to determine. Using this method differences between East and West were discovered in detail and functional correlations between endogenous and exogenous variables were investigated. This analysis unveiled significant differences of labor supply behavior among East and West Germany.

key words: female labor supply, nonparametric regression, GAM, Integration estimator

# 1 Einleitung

Für die Arbeitsmarktforschung stellt die Schätzung des Arbeitsangebots ein zentrales Aufgabenfeld dar. Wie beeinflussen verschiedene Variablen die Entscheidung des Individuums, sich auf den Arbeitsmarkt zu begeben und seine Arbeitskraft anzubieten? Gibt es speziell in Deutschland wenige Jahre nach dem politischen Umschwung im Osten Differenzen im Vergleich der Bewohner zwischen Ost und West, die aus den Erfahrungen und Verhaltensweisen unter verschiedenen Wirtschaftssystemen herrühren?

Die interessanteste zu untersuchende Gruppe ist die der verheirateten Frauen, da sie die größte Variabilität in Verhalten aufweisen, bedingt vor allem durch die hohen häuslichen und familiären Belastungen. Für die Bundesrepublik liegen relativ wenige Analysen zum Arbeitsangebot vor. Die geringe Zahl der Arbeiten zu Deutschland liegt u.a. an der Schwierigkeit, Individualdatensätze zur Verfügung gestellt zu bekommen. Erst in jüngerer Zeit erschienen einige Arbeitsangebotsuntersuchungen, die auf Daten des Sozio-ökonomische Panels (SOEP) des DIW basieren (siehe z.B. Merz, 1990; Strøm und Wagenhals, 1991; Stobernack, 1991).

Ökonometrische Untersuchungen zum Arbeitsangebot wurden bis Mitte der siebziger Jahre im wesentlichen mit OLS-Schätzungen durchgeführt. Erklärende Variablen waren meist der Lohnsatz, das Haushalteinkommen einer Periode und die Arbeitslosenquote als Konjunkturindikator. Mit der zweiten Generation von Studien wurden die methodischen Aspekte mehr in den Mittelpunkt gerückt<sup>1</sup>. So entstanden eine Reihe ökonometrischer Techniken wie Logit oder Probit für Binary-Choice-Modelle<sup>2</sup>. Desweiteren sollten recht viele Einflußparameter identifiziert werden. Die Arbeiten unterschieden nun auch nach den beiden wesentlichen Dimensionen des Arbeitsangebots, der Partizipationswahrscheinlichkeit auf der einen Seite und den angebotenen Stunden am Arbeitsmarkt als Antwortvariable auf der anderen<sup>3</sup>.

Die oben erwähnten Verfahren erfordern jedoch für jede erklärende Variable Kenntnisse über den funktionalen Typ des Zusammenhangs mit der Antwortvariable. Dies kann mit der Verwendung nichtparametrischer Regressionsmethoden verhindert werden. Diese wurden seit Mitte der 60er Jahre entwickelt. Da sie aber relativ rechen-

---

<sup>1</sup>siehe Berndt (1990), S. 595 und 633f.

<sup>2</sup>eine Darstellung dieser Techniken erfolgt z.B. in Ronning (1991), eine Anwendung zu diesem Thema in Merz (1987)

<sup>3</sup>vgl. Heckman (1993)

intensiv sind, konnte eine sinnvolle Anwendung erst in den letzten Jahren geschehen. Ergebnis ist nicht mehr eine analytisch beschriebene Funktion, sondern oft nur eine grafische Darstellung. Da diese Verfahren auf dem Prinzip der lokalen Mittelung beruhen, verschlechtert sich die Regression mit wachsender Dimension und der damit verbundenen Datenausdünnung dramatisch („Fluch der Dimensionalität“). Dieses Problem wird mit der Verwendung Additiver Modelle (AM, siehe Hastie und Tibshirani (1991)) gelöst. Darin wird das Regressionsproblem auf den eindimensionalen Fall durch die Annahme der Separierbarkeit der verschiedenen Einflußfaktoren reduziert.

Dieser Zielrichtung verpflichtet, sind in den achtziger Jahren verstärkt semiparametrische Modelle entwickelt worden. Sie zeichnen sich dadurch aus, daß sie parametrische und nichtparametrische Methoden in sich vereinen. In diese Kategorie fallen vor allem die Single Index Modelle (SIM, siehe Härdle und Turlach, 1992) und die Generalisierten Additiven Modelle (GAM, siehe Hastie und Tibshirani, 1991). Mit diesen sensiblen Methoden sind heute neue Möglichkeiten der Schätzung ökonomischer Beziehungen gegeben und die Überprüfung der Art von Zusammenhängen kann erfolgen.

Trotz Verbesserungen in den letzten Jahren ist das Erwerbsleben verheirateter Frauen noch stark von Diskontinuitäten geprägt. Sie tragen die Hauptlast der familiären Pflichten und insbesondere Kinder führen zu geringerer Mobilität und Zeitdisponibilität. Gerade der deutsche Arbeitsmarkt stellt aber sehr hohe Anforderungen an Verfügbarkeit, Mobilität und aktuellem Humankapital. So sind Familie und Beruf noch immer schwer miteinander zu vereinbaren. Außerdem lassen die deutschen Löhne oftmals einen Verdiener in der Familie ausreichen, dies ist im Normalfall der Mann. Einer der wichtigsten Indikatoren der deskriptiven Statistik, die Erwerbsquote, bestätigt dies. So lag die Frauenerwerbsquote der Bundesrepublik 1988 (54.4 %) deutlich unter dem OECD-Durchschnitt (59.1 %)⁴.

Ganz anders stellte sich die Situation in der DDR dar. Auch hier trugen sicherlich die Frauen die Hauptlast der häuslichen Arbeiten. Doch lag die Erwerbsquote laut Statistischem Jahrbuch der DDR von 1990 im Jahr 1989 bei rund 82 %. Hier trugen in großem Umfang ein entsprechendes gesellschaftliches Klima und die Bereitstellung preiswerter öffentlicher Kindereinrichtungen zu dieser hohen Zahl bei. So war zwar die Erwerbsquote in den Jahren nach der politischen Wende in der DDR stark gefallen, lag jedoch noch immer weit über der Erwerbsbeteiligung im Westen Deutschlands. Auch

---

⁴siehe Franz (1991), S.26

spricht die derzeitige Höhe der ostdeutschen Löhne oft für die Notwendigkeit zweier Verdienner in der Familie. Der hohe gesellschaftliche und soziale Wert der Erwerbsarbeit, die hohe Weiterbildungsbereitschaft und die bereits aus der DDR bekannte Doppelbelastung von Familie und Beruf lassen nicht erwarten, daß die Frauen aus den neuen Bundesländern dem Arbeitsmarkt resigniert den Rücken kehren (Holst und Schupp, 1991 u. 1994). So ist auch ein großer Teil des Rückganges der ostdeutschen Frauenerwerbsquote mit den umfangreichen Vorruhestandsregelungen für ältere Arbeitnehmerinnen zu erklären. Dies alles deutet schon an, daß die Verhaltensmuster zwischen Ost und West starke Differenzen aufweisen.

Ziel der Analyse war es, vermutete noch bestehende Unterschiede im Erwerbsverhalten zwischen den verheirateten Frauen Ost- und Westdeutschlands zu ermitteln. Für deren Herausarbeitung wurde ein semiparametrisches Regressionsmodell verwendet. Dafür sprach auch, daß es so gut wie keine Erfahrungen gab, ob die funktionalen Zusammenhänge zwischen Einflußfaktoren und Erwerbsbeteiligung im Osten Deutschlands genauso anzusetzen sind.

Die Arbeit ist wie folgt gegliedert. Die nächsten beiden Abschnitte befassen sich mit den zur Verfügung stehenden Daten und den benutzten Variablen. Im vierten Kapitel werden theoretische Ausführungen zum verwendeten Regressionsmodell und im fünften zum Schätzverfahren gemacht. In Abschnitt 6 erfolgt die Präsentation der Schätzergebnisse und im letzten Kapitel eine kurze Zusammenfassung.

## 2 Daten

Für eine Analyse der Frauenerwerbstätigkeit sind umfangreiche Daten über die einbezogenen Personen wie Alter, Geschlecht, Bildung, Berufserfahrung, Einkommen wie auch den familiären Hintergrund (Haushalteinkommen, Kinderzahl usw.) nötig. Da diese Informationen aus datenschutzrechtlichen Gründen nicht von der amtlichen Statistik geliefert werden können, schließt die Forschung selbst diese Lücke.

Hier ist in erster Linie das Sozio-ökonomische Panel (SOEP) in Verantwortung des Deutschen Institutes für Wirtschaftsforschung (DIW) zu nennen. Das SOEP wird seit 1984 in der damaligen Bundesrepublik mit jährlich einer umfangreichen Befragungswelle im Frühjahr geführt. In der ersten Welle wurden 12245 Personen über 16 Jahren in 5921 Haushalten befragt. Um Längsschnittstudien zu ermöglichen, werden dieselben Personen verfolgt. Auch Kinder befragter Haushalte werden ab dem Alter von 16 Jah-

ren befragt und weiter verfolgt. Bereits im Juni 1990 wurde die erste Befragung noch in der DDR durchgeführt, so daß umfangreiche Informationen auch zu Ostdeutschland zu Beginn des Transformationsprozesses gesammelt werden konnten. In die vorliegende Untersuchung sind nur verheiratete Frauen deutscher Nationalität einbezogen worden, da bei Frauen anderer Nationalitäten ein anderes Erwerbsverhalten erwartet wird.

Die SOEP-Befragung unterteilt sich in einen Haushaltfragebogen und die Personenfragebögen für jede Person im Alter von über 16 Jahren. In jeder Welle werden Fragen zur Familiensituation, Erwerbsleben, Bildung, Einkommen und Eigentum, Freizeit, Wohnen und Lebenszufriedenheit gestellt. Neben den Kernfragen gibt es jedes Jahr einen Erhebungsschwerpunkt mit tiefergehenden Fragen zu einem Thema<sup>5</sup>. Da das SOEP ein Mehr-Zweck-Panel ist, sind gewünschte spezielle Informationen teilweise nicht enthalten oder nur indirekt aus den Angaben zu ermitteln.

Für die Analyse wurden zwei Datensätze angelegt, getrennt nach ost- und westdeutschen Frauen. Es liegen die Daten der Befragungswelle 10 von 1993 zugrunde. Zu diesem Zeitpunkt erfaßte das SOEP im ganzen rund 13200 Personen über 16 Jahren. Aufgrund vieler noch in Ausbildung bzw. bereits im Ruhestand befindlicher Frauen wurde die Altersgrenze für die Analyse relativ eng gewählt<sup>6</sup>. Es wurden nur Frauen im Alter von über 25 Jahren bis 60 Jahren berücksichtigt, alle sind verheiratet. Auch innerhalb der Altersgrenze wurden alle Studentinnen, Auszubildende und Rentnerinnen ausgeschlossen, da sie dem Arbeitsmarkt gar nicht zur Verfügung stehen konnten. Unvollständige Datensätze wurden ebenfalls ausgeschlossen, der Verlust betrug etwa 13 % der noch vorhandenen Datensätze. Da alle ermittelten Variablen in ähnlichem (geringem) Umfang an den „Missings“ beteiligt waren, ist die Gefahr einer systematischen Verzerrung der Stichprobe durch deren Ausschluß als sehr gering einzuschätzen.

Es konnten 1283 Frauen aus dem Westen und 958 Frauen aus dem Osten des Landes in die Untersuchung einbezogen werden. Die Anteile der erwerbstätigen Frauen stimmten in beiden Stichproben weitgehend mit der offiziellen Erwerbsquote überein (Ost: 65%, West: 54%). Die Analysen wurden für jeden Datensatz getrennt durchgeführt.

---

<sup>5</sup>siehe Schupp und Wagner (1995)

<sup>6</sup>Zum Erwerbspotential zählen alle Personen im Alter von 15 bis 65 Jahre.

### 3 Variablen

Es soll untersucht werden, wie die Entscheidung für eine Erwerbstätigkeit von verschiedenen Faktoren beeinflusst wird. Die abhängige Variable oder der Regressand  $Y$  für das Schätzmodell ist demnach der Erwerbstätigkeitsstatus. Dieser hat nur zwei Ausprägungen, es handelt sich um ein Binary-Choice-Modell mit

$$Y = \begin{cases} 0 & \text{Frau ist nicht erwerbstätig} \\ 1 & \text{Frau ist erwerbstätig (auch in Teilzeitbeschäftigung)} \end{cases}$$

Die erklärenden oder unabhängigen Variablen  $X = (x_1, \dots, x_8)$  sind auf der Basis der Erkenntnisse der deskriptiven Statistik zu diesem Thema, der Humankapitaltheorie und anderer relevanter ökonomischer Theorien ausgewählt worden. Natürlich spielte die Beschaffbarkeit der Daten und die Quantifizierbarkeit der Merkmale ebenfalls eine Rolle. Im einzelnen sind folgende Einflußfaktoren einbezogen worden:

- Alter der Frau. Junge Frauen haben eine sehr hohe Erwerbsquote, diese sinkt mit der Geburt der Kinder. Sind die Kinder älter oder erwachsen, kehren die Frauen häufig wieder in die Berufstätigkeit zurück. Man spricht in diesem Zusammenhang auch vom Drei-Phasen-Modell des Erwerbsverhaltens<sup>7</sup>. In den letzten Jahren vor dem Rentenalter sinkt der Anteil der erwerbstätigen Frauen stark ab (siehe Tabelle 1, hier wird auch der Einfluß der ostdeutschen Frauen auf die gesamtdeutsche Frauenerwerbsquote deutlich).

Altersklasse von ... Jahren bis unter ... Jahren	April 1990 alte Bundesländer	April 1991 Deutschland
15 – 20	39.5 %	49.7 %
20 – 25	63.7 %	70.1 %
25 – 30	62.3 %	68.7 %
30 – 35	60.0 %	67.4 %
35 – 40	63.1 %	71.2 %
40 – 45	65.5 %	72.1 %
45 – 50	62.3 %	69.4 %
50 – 55	53.3 %	61.8 %
55 – 60	39.2 %	39.3 %
60 – 65	11.0 %	9.5 %
Mittel der 15 – 65-jährigen	53.9 %	59.7 %

Tabelle 1: Die Erwerbsquoten verheirateter Frauen der alten Bundesländer (4/90) und Gesamtdeutschlands (4/91). Quelle: Statistisches Jahrbuch (1992, 1993).

<sup>7</sup>Myrdal, A. und Klein, V. (1960) *Die Doppelrolle der Frau in Familie und Beruf*, Kiepenheuer und Witsch, Köln.



- Stundenlohn. Der Lohn ist als Preis der Arbeit der wichtigste Indikator in der Arbeitsmarktökonomik. Sobald der individuelle Anspruchslohn einer Person den Lohn, den sie auf dem Arbeitsmarkt erzielen kann, übersteigt, geht laut Theorie das Individuum arbeiten.
- Bildungsstand. Laut Humankapitaltheorie sind Investition in die eigene Bildung mit Erwartungen über erhöhte Rückflüsse aus der Beschäftigung am Arbeitsmarkt verbunden. Für die Analyse wurden die Bildungsabschlüsse pauschal mit Jahren in Bildung bewertet (z.B. ein Hochschuldiplom mit 5 Jahren), so daß ein quantifiziertes quasistetiges Merkmal entstand.
- Miete. Da die Wohnkosten meist den größten Ausgabenposten eines Haushaltes darstellen, soll die Belastung des Wohnens in Höhe der Kaltmiete bzw. die Tilgungsrate bei Eigentumserwerb in die Schätzung einbezogen werden.
- Arbeitslosenquote des Bundeslandes. Die Arbeitslosenquote ist als regionaler Konjunkturindikator in die Analyse einbezogen worden.
- Anzahl der Kinder. Für die Entscheidung der Frauen für oder gegen eine Erwerbstätigkeit spielt das Vorhandensein bzw. die Zahl der Kinder eine entscheidende Rolle. Im SOEP sind Personen unter 16 Jahren als Kinder erfaßt, diese Klassifizierung wurde beibehalten.
- Nettolohn des Ehemannes. Wieweit spielt die finanzielle Absicherung durch den Ehepartner eine Rolle?
- Sonstiges verfügbares Haushaltseinkommen. In dieser Variable sind Rentenzahlungen, Sozialtransfers, Kapitaleinkommen, Mieteinnahmen usw. enthalten. Sie kann auch negativ sein, da hier Auszahlungen für Alimente, Stützung nicht im Haushalt lebender Angehöriger und geschiedener Ehepartner vorher abgezogen werden.

Wichtig für die Interpretation der Ergebnisse ist, daß durch die Kategorisierung im SOEP unfreiwillig Arbeitslose als nicht erwerbstätig gelten, obwohl sie Arbeit anbieten. So ist deren Antwortvariable  $Y = 0$ . D.h., als erwerbstätig ( $Y = 1$ ) zählen nur die Frauen, die sich auf den Arbeitsmarkt begeben haben **und** erfolgreich waren bei der Arbeitssuche.

## 4 Modell

Für die Erläuterung des verwendeten Schätzmodells, einem Generalisierten Additiven Modell (GAM), gehen wir bei der Betrachtung von einem multivariaten linearen Regressionsmodell aus. Dieses parametrische Standardwerkzeug der Statistik hat die Form

$$Y = X^T \beta + \varepsilon = c + X_1 \beta_1 + \dots + X_p \beta_p + \varepsilon, \quad (1)$$

mit Fehlerterm  $\varepsilon$  unabhängig von  $X_j$ , einen Stichprobenumfang  $i = 1, \dots, n$  und den Ausprägungen des  $i$ -ten Objektes  $y_i, x_{i1}, \dots, x_{ip}$  der  $p$  betrachteten Merkmale. Die  $\varepsilon_i$  sind unabhängig und identisch verteilt mit

$$E \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \text{Cov} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \text{Cov} \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}.$$

Die Beliebtheit dieses Modells gründet sich neben seiner unkomplizierten Handhabung darauf, daß die Effekte der einzelnen Merkmale auf die Antwortvariable additiv verbunden sind unter der Voraussetzung fehlender Multikollinearitäten. So ist eine leichte und nach Einflüssen separierte Interpretation möglich.

Ein wesentlicher Nachteil des Regressionsmodells (1) ist die Ermittlung von nur linearen Zusammenhängen. Eine Verallgemeinerung führt zum additiven Modell (AM)

$$Y = c + \sum_{j=1}^p g_j(X_j) + \varepsilon. \quad (2)$$

Die Parameter  $\beta_j$  wurden ersetzt durch Funktionen  $g_j$ . Das lineare Regressionsmodell (1) stellt sich als Spezialfall additiver Modelle (2) dar. Wie in (1) gelten auch hier die Eigenschaften für  $\varepsilon$  wie oben.

Die Regressionsfunktion lautet

$$m(x) = E(Y|X = x) = c + g_1(x_1) + \dots + g_p(x_p) \quad (3)$$

mit  $X = (x_1, \dots, x_p)$  als Vektor der erklärenden Variablen,  $c$  ist eine Konstante,  $\{g_j(\bullet)\}_{j=1}^p$  ist eine Menge unbekannter Funktionen, die  $E_{X_j} g(X_j) = 0$  erfüllen und  $x = (x_1, \dots, x_p)$  sind die realisierten Werte. Jede der Funktionen  $g_j$  ist univariat und beinhaltet nur relative Einflüsse auf  $Y$ . Jede der univariaten Funktionen  $g_j(x_j)$  wird separat nichtparametrisch geschätzt, in dieser Analyse mit Hilfe von Kerndichteschätzungen (siehe Härdle, 1990). Somit ist das in der Einleitung angesprochene

Problem des „curse of dimensionality“ gelöst, da eine  $p$ -dimensionale Schätzung asymptotisch reduziert worden ist auf  $p$  eindimensionale Schätzungen.

Nach Ermittlung von  $g_j$  aus den Daten kann jede der  $p$  Funktionen einzeln grafisch dargestellt werden. Diese Grafiken erlauben die Interpretation der Effekte marginaler Änderungen einer Variable  $X_j$  auf die Antwortvariable  $Y$ . Dadurch ist eine übersichtliche Darstellung auch mehrerer Einflußfaktoren ( $p \geq 3$ ) möglich.

Für jede Person führt die Summe der Einflußfaktoren der Angebotsentscheidung zu einem individuellen unbeobachtbaren Wert der abhängigen Variable  $Y^*$ , die als „latente“ Variable bezeichnet wird:

$$Y^* = c + \sum_{j=1}^p g_j(X_j) - \varepsilon. \quad (4)$$

$\varepsilon$  ist die stochastische Restkomponente, die die nicht erfaßten Einflußgrößen berücksichtigt (Annahme:  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$  unabhängig von  $X$ ). Beobachtet werden kann nur die Realisierung am Arbeitsmarkt: bietet die Person ihre Arbeitskraft an oder nicht. Die beobachtete Antwortvariable  $Y$  hat eine dichotome Ausprägung. Überschreitet  $Y^*$  einen bestimmten Schwellenwert  $\alpha$ , entscheidet sich die Person für eine Arbeitsaufnahme ( $Y^* \geq \alpha \Rightarrow Y = 1$ ), sonst ist  $Y = 0$  (keine Arbeitsaufnahme).

Mit dem Auftreten der 0-1-Antwortvariablen kann das Modell (3) nicht mehr richtig sein, da nicht sichergestellt ist, daß  $E(Y|X = x)$  zwischen 0 und 1 liegt.  $E(Y|X = x)$  ist in diesem Kontext die Teilnahmewahrscheinlichkeit der Frauen am Arbeitsmarkt. Mit dem Vorschalten einer sogenannten „Linkfunktion“  $G$ , die einen Definitionsbereich zwischen 0 und 1 hat, wird das Problem gelöst. Das neue Modell lautet somit

$$m(x) = E(Y|X = x) = G\left(c + \sum_{j=1}^p g_j(X_j)\right). \quad (5)$$

Unter der bereits oben genannten Annahme ( $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ ) leitet sich für die Linkfunktion die Verteilungsfunktion der Normalverteilung her<sup>8</sup>.

Modell (5) wird als Generalisiertes Additives Modell (GAM) bezeichnet. Es vereint die nichtparametrischen Komponenten aus Modell (3) mit einer parametrischen (da als bekannt angenommenen) Funktion  $G$  zu einem semiparametrischen Modell<sup>9</sup>.

Abbildung 1 veranschaulicht die Einordnung der GAM in ein System von Regressionsmodellen.

<sup>8</sup>Herleitung siehe z.B. Ronning (1991), S. 8-10

<sup>9</sup>siehe Hastie und Tibshirani (1991)

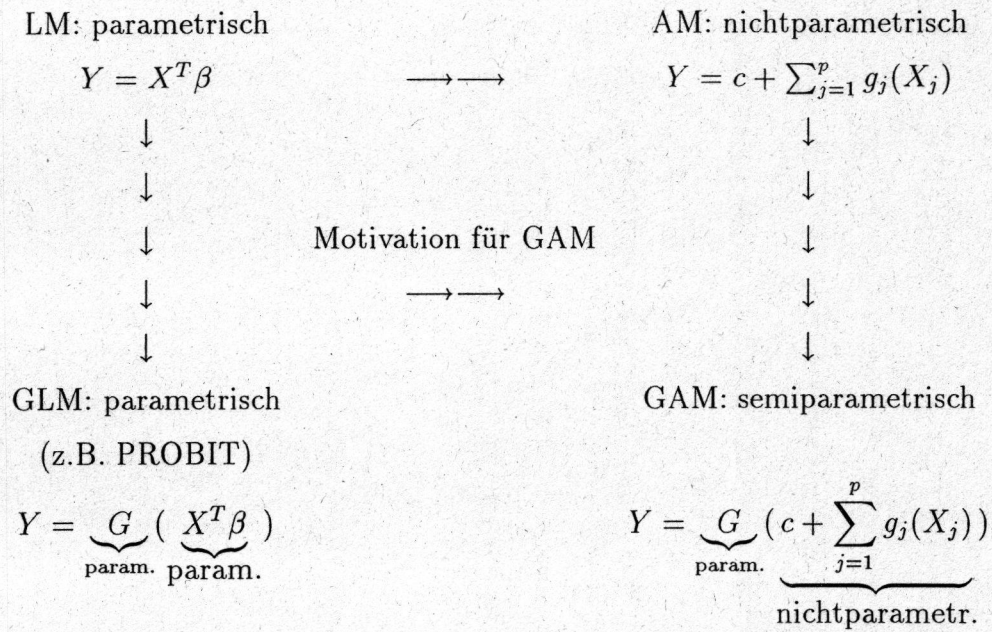


Abbildung 1: Einordnung der GAM.

Eine etwas andere Systematik, geordnet nach bekannten und nicht bekannten Bestandteilen, zeigt Tabelle 2.

Link-Fkt.	bekannt	unbekannt
Komponenten		
linear	GLM: $Y = G(X^T \beta)$	SIM: $Y = g(X^T \beta)$
nichtparametrisch	GAM: $Y = G(c + \sum g_j(X_j))$	?

Tabelle 2: Parametrische und semiparametrische Regressionsmodelle.

Wenn  $\Phi(\bullet)$  die Verteilungsfunktion der Normalverteilung ist, so lautet das für die Schätzung benutzte Modell:

$$m(x) = E(Y|X = x) = \Phi\left(c + \sum_{j=1}^8 g_j(x_j)\right). \quad (6)$$

## 5 Schätzverfahren

Das bisher einzige bekannte Schätzverfahren im GAM war das Backfittingverfahren, ein Iterationsverfahren mit Anfangswert, Zyklus und Abbruchkriterium. Für ein Modell der Form  $E(Y|X) = c + \sum g_j(X_j)$ <sup>10</sup> wurden in einem ersten Schritt die Anfangswerte festgelegt:  $\hat{g}_1^{(0)}, \hat{g}_2^{(0)}, \dots, \hat{g}_p^{(0)}$ . Im zweiten Schritt erfolgte die Schätzung der  $g_j$ , wobei wegen  $E(g_j(X_j)) = 0 \forall j$  folgte:  $\hat{c} = E(Y|X)$ . Die Schätzung beruhte auf dem Prinzip der partiellen Residuen (Gauss-Seidel-Algorithmus):

$$\begin{aligned} \hat{g}_1^{(1)} &= Y - \hat{c} - \hat{g}_2^{(0)} - \dots - \hat{g}_p^{(0)}, \\ \hat{g}_2^{(1)} &= Y - \hat{c} - \hat{g}_1^{(1)} - \hat{g}_3^{(0)} - \dots - \hat{g}_p^{(0)}, \\ &\vdots \\ \hat{g}_p^{(1)} &= Y - \hat{c} - \hat{g}_1^{(1)} - \dots - \hat{g}_{p-1}^{(1)}. \end{aligned} \tag{7}$$

Dieser Zyklus wurde nun immer wieder wiederholt, bis das Abbruchkriterium erfüllt wurde<sup>11</sup>:  $|\hat{g}_j^{(l)} - \hat{g}_j^{(l-1)}| < \alpha$ .

Einige Eigenschaften dieses Verfahrens wirken sich jedoch nachteilig aus. So sind die Ergebnisse abhängig von der Reihenfolge der Schätzung sowie von den gewählten Anfangswerten. Es ist nicht bekannt, ob beim Abbruch Konvergenz überhaupt schon begonnen hat. Das Abbruchkriterium ist intuitiv festzulegen. Die analytischen Eigenschaften des Schätzers sind in der Praxis nicht bestimmbar<sup>12</sup>.

Seit etwa einem Jahr sind neue Ansätze, in dieser Modellform zu schätzen, bekannt. Der Integrationsschätzer ist ein analytischer Schätzer und basiert auf einer „Integrationsidee“ von Linton und Nielsen (1995).

Die Integrationsidee ist die folgende<sup>13</sup>: Ausgangspunkt ist ein additives Modell der Form (3) wie in Abschnitt 4 beschrieben<sup>14</sup>. Die Funktionen  $g_j(\bullet)$  haben Erwartungswert von 0, d.h.  $\int g_j(\omega) f_j(\omega) d\omega = 0$ , wobei  $f_j$  die marginale Dichte von  $x_j$  ist. Die

<sup>10</sup>Die Verwendung einer Linkfunktion eines GAM ist für die nichtparametrische Schätzung der  $g_j$  ohne Bedeutung. Darum reicht die Betrachtung im AM aus.

<sup>11</sup>ausführliche Beschreibung dieses Verfahren z.B. in Hastie und Tibshirani (1991)

<sup>12</sup>Die Eigenschaften des Verfahrens sind intensiv dargestellt und diskutiert in Härdle und Hall (1993).

<sup>13</sup>vgl. Chen, Härdle, Linton und Severance-Lossin (1995)

<sup>14</sup>Es gilt wie oben: die Betrachtung im AM reicht aus.

Regressionsfunktion lautet

$$m(x_1, \dots, x_p) = c + \sum_{j=1}^p g_j(x_j).$$

Man kann die jeweils nicht interessierenden Dimensionen mit Hilfe ihrer gemeinsamen marginalen Dichte  $f_{-j}$  herausintegrieren. Denn es gilt

$$g_j(x) + c = \int m(x_1, \dots, x, \dots, x_p) f_{-j}(x_1, \dots, x_p) \prod_{s \neq j} dx_s, \quad (8)$$

wobei  $f_{-j}$  die gemeinsame Dichte von  $X_{i1}, \dots, X_{i(j-1)}, X_{i(j+1)}, \dots, X_{ip}$  ist.

Da (8) ein Erwartungswert ist, kann  $\hat{g}_j(x)$  mit dem Durchschnitt für endlichen Stichprobenumfang  $n$  geschätzt werden:  $\frac{1}{n} \sum_{i=1}^n m(\bullet)$ . Mit dem Nadaraya-Watson-Kernschätzer wird die Funktion  $m(\bullet)$  bestimmt. Dazu wird der Durchschnitt über die Beobachtungen gebildet, die mit dem Schätzer

$$\begin{aligned} \hat{g}_j(x) &= \frac{1}{n} \sum_{i=1}^n \hat{m}(X_{i1}, \dots, X_{i(j-1)}, x, X_{i(j+1)}, \dots, X_{ip}) \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{\sum_{l=1}^n [\prod_{s \neq j} L_{h'_n}(X_{ls} - X_{is})] K_{h_n}(X_{lj} - x) Y_l}{\sum_{t=1}^n [\prod_{s \neq j} L_{h'_n}(X_{ts} - X_{is})] K_{h_n}(X_{tj} - x)} \right] \end{aligned} \quad (9)$$

ermittelt werden, wobei  $h_n$  die Bandweite des Kerns  $K$  der uns interessierenden  $j$ -ten Dimension und  $h'_n$  die Bandweiten der Kerne  $L$  der anderen  $s \neq j$  Dimensionen sind.

Ein Nachteil des Integrationsschätzers ist, daß mit der Wahl mindestens zweier Bandweiten ( $h_n$  und  $h'_n$ ) mehr Parameter zu wählen sind als im Backfitting. Jedoch überwiegen die Vorteile des Verfahrens. Es sind analytische Eigenschaften wie die asymptotische Verteilung von Bias, Varianz und Konvergenzrate bestimmbar. Da auch die Ableitungen der additiven Funktionen schätzbar sind, ist es möglich, Elastizitäten und Substitutionsraten zu bestimmen<sup>15</sup>. Die grafischen Ergebnisse der Schätzung in Abschnitt 6 konnten mit Konfidenzintervallen angegeben werden.

Die Konsistenz des Integrationsschätzers ist bewiesen, d.h. es ist gezeigt, daß für  $n \rightarrow \infty$  Bias und Varianz Null werden<sup>16</sup>:  $\hat{g}(x_i) - g(x_i) \xrightarrow{p} 0 \quad \forall i, \quad nh \rightarrow \infty$ . Der Schätzer ist asymptotisch normalverteilt. In ersten Untersuchungen mit simulierten Daten wurden Varianz und Bias bestimmt, um die Qualität der beiden Schätzer zu vergleichen (Kempe, 1995). Es stellte sich heraus, daß das neue Verfahren mindestens ebensogute Resultate lieferte wie das Backfitting, oft auch überlegen war. Während

<sup>15</sup>siehe Severance-Lossin und Sperlich (1995)

<sup>16</sup>siehe Chen, Härdle, Linton und Severance-Lossin (1995)

der alte Schätzer die gesamte Regressionsfunktion  $\hat{m}$  oft besser schätzte<sup>17</sup>, waren Varianz und Bias in den einzelnen additiven Funktionen  $\hat{g}_j$  meist beim Integrationsverfahren kleiner. Dies ist neben den theoretischen Vorteilen ein weiterer Vorzug des neuen Schätzers. Für die separate Interpretation jeder Variablen gilt das Interesse den Schätzergebnissen der einzelnen additiven Funktionen. Als problematisch ist jedoch die Bandweitenwahl anzuführen, da die Gefahr des Über- oder Unterglätzens der  $\hat{g}_j$  gegeben ist. Desweiteren waren die Schätzungen am Rand schlecht, da die Regression bei geringen Datenmengen wenig robust ist. Bei der Analyse hatte das sehr weite Konfidenzintervalle zur Folge, so daß Ausreißerwerte nicht in die Analyse einbezogen worden sind.

Es wurde aufgrund des Problems der Selektionsverzerrung ein mehrstufiges Schätzverfahren notwendig. Für die Personen, die keine Arbeit realisieren, kann kein Lohn beobachtet werden. Der Lohn ist jedoch eine zentrale Variable der Arbeitsangebotsentscheidung.

Die Schätzung bleibt bei der Einbeziehung nur des arbeitenden Teils der Stichprobe nicht mehr konsistent. Diese Untergruppe stellt keine zufällige Auswahl mehr dar, sie ist vielmehr systematisch. So sind die ermittelten Schätzergebnisse verzerrt bzw. nur für diese Teilgruppe gültig. Heckman (1979) hat nun eine Korrektur der Verzerrung vorgeschlagen, die ein dreistufiges Verfahren für diese Analyse notwendig gemacht hat:

1. Reduzierte GAM-Schätzung zur Ermittlung einer Selektionskorrigierenden Variable  $\hat{\lambda}$ .
2. Selektionskorrigierte Schätzung der Variable Lohn mit dem OLS-Verfahren unter Einbeziehung von  $\hat{\lambda}$ .
3. Die eigentliche GAM-Schätzung.

In einer ersten Stufe wird eine „Selektionsvariable“  $\lambda$  geschätzt, die die Verzerrung wieder korrigiert. Dieses  $\hat{\lambda}$  enthält die Wahrscheinlichkeit, zur Gruppe der Arbeitenden ( $Y = 1$ ) zu gehören<sup>18</sup>.

$$\hat{\lambda} = f(\text{Kinderzahl, Bildung, AL-Quote, potent. Berufserfahrung, Miete, Eink. Ehemann, sonst. Haushaltseink.})$$

<sup>17</sup>Das ist dadurch zu erklären, da das Backfittingverfahren auf die Residuen schätzte, vgl. Formel (7) auf Seite 11 oben.

<sup>18</sup>Eine Erklärung und Herleitung findet sich in Franz (1991), S. 80f.

Diese geschätzte Selektionsvariable bezieht die Informationen anderer vorhandener Variablen aller Individuen ein, um die Wahrscheinlichkeit, zur Gruppe der Erwerbstätigen zu gehören, zu ermitteln.

Mit Hilfe dieses  $\hat{\lambda}$  und anderer Variablen kann nun in einer zweiten Stufe für alle in die Untersuchung einbezogenen Personen, also auch die nicht erwerbstätigen, ein individueller Lohn geschätzt werden, den diese Personen theoretisch verdienen könnten (aber nicht realisiert haben).

$$\hat{w} = f(\hat{\lambda}, \text{Bildung, AL-Quote, beobacht. Lohn, potent. Berufserfahrung})$$

Mit dem hier ermittelten Lohn  $\hat{w}$  wird für alle Personen weitergerechnet<sup>19</sup>.

Zuletzt erfolgt die eigentliche GAM-Schätzung. Jetzt liegen für alle Individuen Lohnangaben vor und die Schätzung der additiven Funktionen  $\hat{g}_j$  kann mit den in Abschnitt 3 beschriebenen Variablen erfolgen.

## 6 Resultate

Die Ergebnisse der Schätzung, die nichtparametrischen Funktionen  $g_j$ , sind in grafischer Form in den Abbildungen 2 - 8 dokumentiert. Die Darstellung erfolgt mit Konfidenzintervallen, deren Punkte verbunden sind<sup>20</sup>. Das verwendete Signifikanzniveau beträgt je Seite 5%<sup>21</sup>. Am Rand wurden jeweils 1-2% der Daten abgeschnitten. Bei zu geringer Datendichte wären die Konfidenzintervalle zu breit.

Jede Abbildung enthält vier Grafiken, oben befinden sich jeweils die nichtparametrischen additiven Funktionen. Zum besseren Ost-West-Vergleich enthält jede Abbildung links die Schätzung für die alten Bundesländer, rechts für die neuen. Unter den additiven Funktionen befindet sich jeweils die Häufigkeitsdichte des betrachteten Merkmals im selben Maßstab, um deutlich zu machen, wo die Masse der Merkmalsausprägungen liegt bzw. wo die Datendichte geringer ist.

Die erste betrachtete Einflußvariable ist der Stundenlohn. Für die Frauen im Westen des Landes ergibt sich ein fast lehrbuchmässiges Bild. Mit zunehmenden zu erwartenden Einkommen ist es für die Frauen auch attraktiv, auf den Arbeitsmarkt zu gehen.

<sup>19</sup>Die potentielle Berufserfahrung ist einfach ermittelt aus dem Alter minus sechs minus Bildung.

<sup>20</sup>Die  $\hat{g}_j(x_j)$  sind asymptotisch normalverteilt.

<sup>21</sup>Durch die Verwendung geschätzter Regressoren (Lohn) liegt eine zusätzliche Unsicherheit vor, die Einfluß auf die Breite der Konfidenzintervalle hat. Diese konnte in den Darstellungen nicht berücksichtigt werden, da sie bisher nicht quantifizierbar ist.



Der Anstieg ist fast linear und mit den engen Konfidenzintervallen als recht sicher anzusehen. Allerdings muss einschränkend darauf aufmerksam gemacht werden, daß in den Abbildungen punktweise Konfidenzintervalle dargestellt sind und keine (breiteren) Konfidenzbänder.

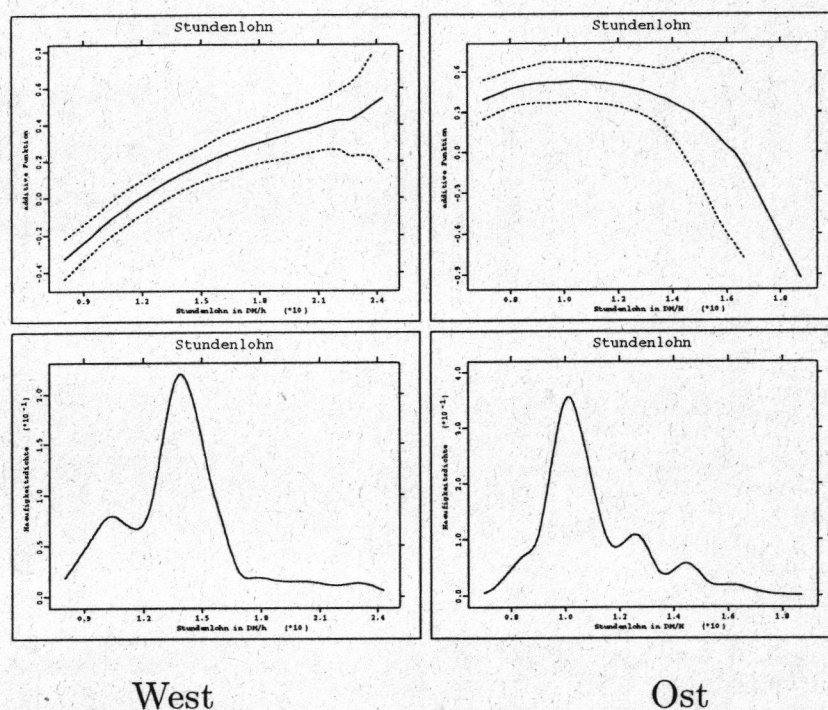


Abbildung 2: Variable *Stundenlohn*.

Dagegen ist das Resultat für die Frauen im Osten doch überraschend. Die Funktion zeigt einen annähernd negativen quadratischen Verlauf. Dabei liegt die Erwerbsneigung bei geringen Löhnen schon sehr hoch und steigt bis zu den mittleren Einkommen nur gering an. Mit weiter steigenden Löhnen sinkt der Anteil der arbeitenden Frauen paradoxerweise stark ab. Eine Erklärung dieses Verlaufs ist mit dem am Ende des Abschnitt 3 erwähnten Erwerbstätigenkonzepts möglich. Die Arbeitslosen wurden zu den Erwerbslosen gezählt. Je höher die zu erwartenden Löhnen sind (aufgrund von Qualifikation, Berufserfahrung usw.), desto schwieriger wird es, einen Arbeitsnachfrager zu finden, der diesen Lohn zu zahlen bereit ist. 1992/93 war der Höhepunkt des wirtschaftlichen Zusammenbruchs in Ostdeutschland mit einer extrem hohen Arbeitslosigkeit und einer Vielzahl arbeitsmarktbegleitender Maßnahmen. So dürfte hier ein zweiter Effekt den „normalen“ steigenden Verlauf überwiegen. Hochqualifizierte und ältere Frauen, die mit einem überdurchschnittlichen Lohn zu rechnen hätten, sind von

Arbeitslosigkeit und Vorruhestand in einem Maße erfaßt, wie es im Westen unbekannt ist. Frauen sind in der Regel von Arbeitslosigkeit und krisenhaften Situationen schwerer betroffen als Männer. Dazu kommt die Entwertung des Humankapitals durch den Systemwechsel.

Relativierend muß aber angefügt werden, daß auch eine Gerade mit Anstieg von Null zwischen die Konfidenzintervalle passen würde. Aufgrund der geringen Datenmenge explodieren gegen Ende die Intervalle, so daß sie auch nicht weiter eingezeichnet wurden. Das Ergebnis bleibt jedoch überraschend, da dies bedeutet, daß mit Veränderung des zu erwartenden Lohnes keine Änderung im Erwerbsverhalten zu beobachten wäre!

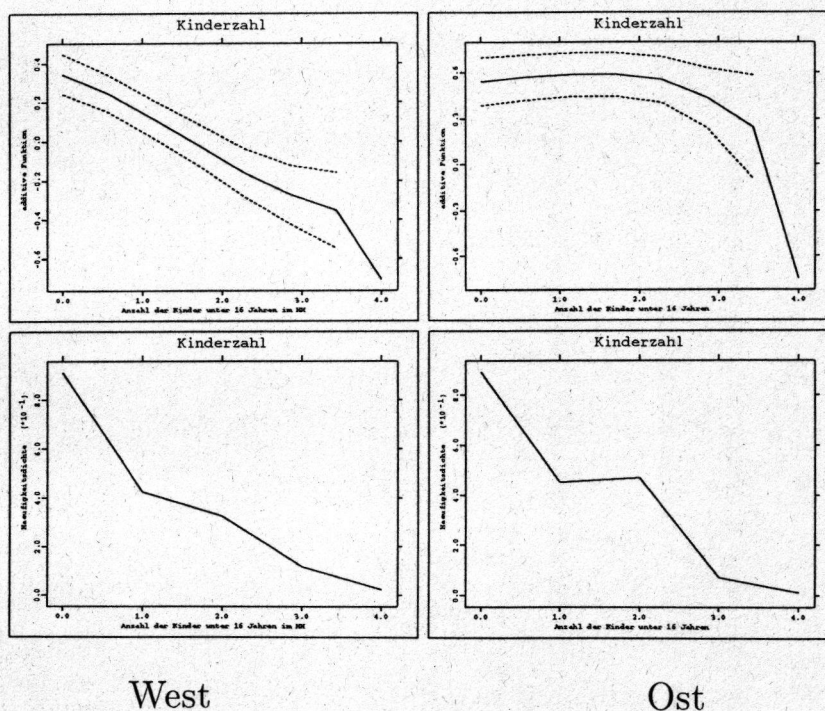


Abbildung 3: Variable Anzahl der Kinder.

Ein weiterer sehr wichtiger Einflußfaktor für die Entscheidung zur Erwerbstätigkeit von Frauen sind die Kinder im Haushalt. Die Kurve für die Frauen in den alten Bundesländern bestätigt bisherige Erkenntnisse, es besteht ein starker negativer linearer Zusammenhang. Kinder sind für die Ausübung einer Berufstätigkeit in einem hocheffizienten Industrieland hinderlich. In dieser Hinsicht ist der Kurvenverlauf für die neuen Bundesländer wiederum erstaunlich verschieden. Bei Frauen ohne bis zu zwei Kindern ist kein signifikanter Zusammenhang zwischen dem Vorhandensein dieser Kinder und

dem Arbeitsangebot zu erkennen. Hier wirken Verhaltensmuster aus der DDR, da dort die Erwerbstätigkeit von Frauen mit Kindern selbstverständlich war. Die Versorgung mit Kinderbetreuungseinrichtungen ist auch Jahre nach der Vereinigung noch wesentlich besser und finanziell günstiger als im Westen des Landes. Ebenso eine Rolle spielt, daß bei ostdeutschen Einkommen ein oder zwei Kinder oft einen zweiten Verdiener nahelegen. Erst mit dem dritten, insbesondere aber mit dem vierten Kind wird es doch sehr schwierig, Kindererziehung und Berufstätigkeit miteinander zu verbinden.

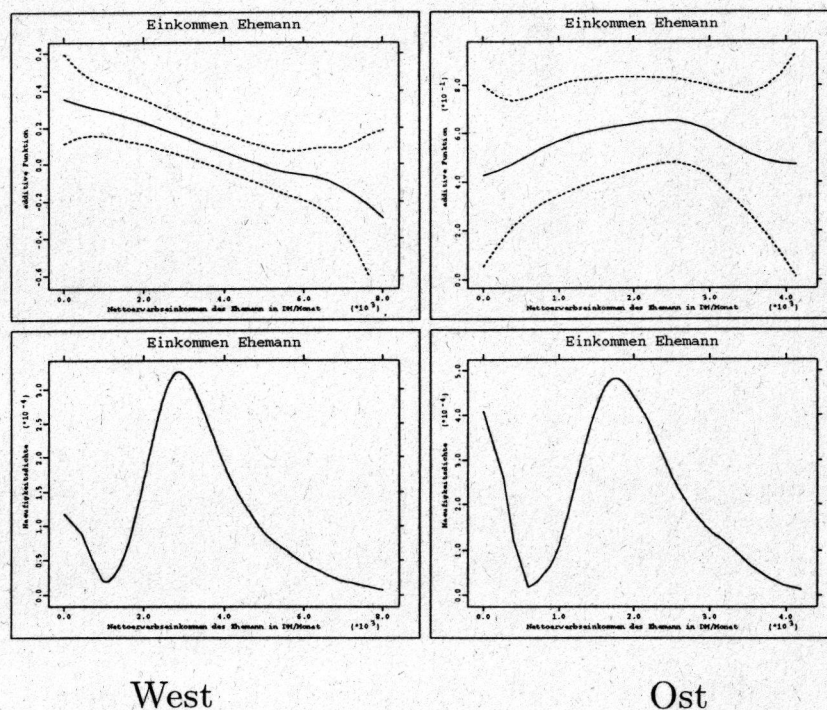


Abbildung 4: Variable *Einkommen des Ehemannes*.

Abbildung 4 illustriert den Zusammenhang zwischen dem Verdienst des Ehepartners und der Neigung, selbst arbeiten zu gehen. Für den Westen Deutschland liegt wiederum ein negativer linearer Funktionsverlauf vor. Dies bedeutet, je mehr der Mann verdient, desto seltener geht die Frau arbeiten. Im Osten ist kein so einheitliches Verhalten zu ermitteln. Jedoch ist im Gebiet der meisten Beobachtungen ein leichter positiver Trend zu erkennen. Dieser ist aber sehr schwach (vgl. Breite der Konfidenzintervalle). Bedingt durch die nachwirkende stärkere Nivellierung der Ost-Gehälter sind die Ehefrauen unabhängig vom Verdienst des Ehepartners erwerbstätig. Da die Partner meist derselben sozialen Schicht und Altersgruppe angehören, entwickeln sich deren Verdienste ähnlich. Bei hohem Einkommen des Ehemannes sinkt aber auch hier die

Erwerbsneigung. Zu beachten ist, daß die Abszisse im Westen Werte zwischen 0 und 8000 DM/Monat Nettoeinkommen enthält, im Osten aber nur bis gut 4000 DM/Monat reicht, ehe das letzte Prozent der Ausreißerwerte erreicht ist.

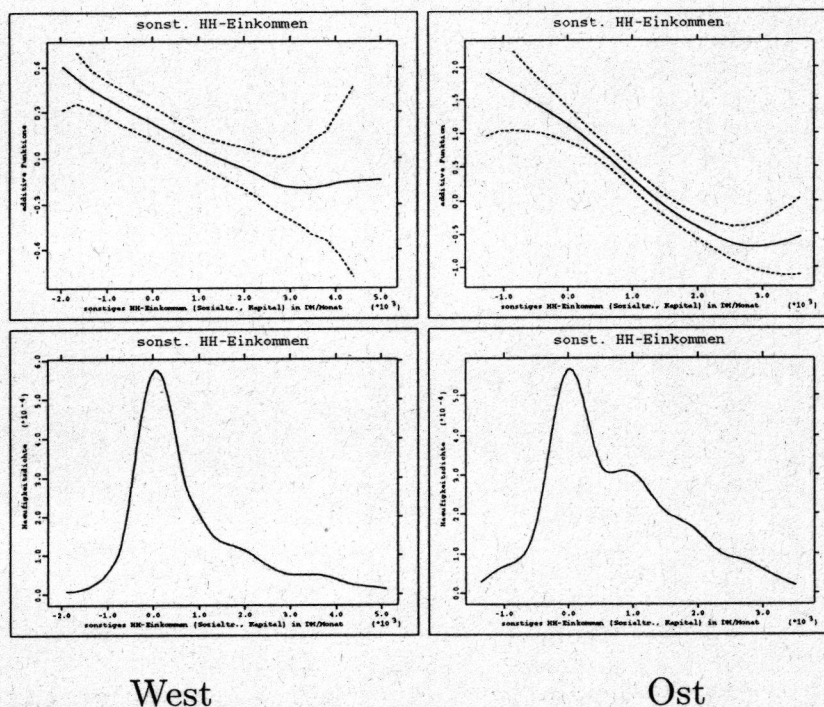


Abbildung 5: Variable *sonst. HH-Einkommen*.

Die Variable sonstiges Haushaltseinkommen (Abbildung 5) besteht vor allem aus Rentenzahlungen, Sozialtransfers wie Kindergeld, Wohngeld, Sozialhilfe sowie den Kapitaleinkommen und Mieteinnahmen. Hier ist es doch überraschend, in beiden Gebieten einen so ähnlichen Zusammenhang beobachten zu können. Sind diese Einnahmen sehr hoch, gehen die Frauen weniger arbeiten. Sind jedoch hohe Belastungen<sup>22</sup> zu verzeichnen, ist die Ehefrau oft erwerbstätig. Aus diesen Bildern ist nicht zu erkennen, ob die Sozialleistungen sich so negativ auf die Erwerbstätigkeitsentscheidung auswirken oder wieviel Verantwortung die Kapitaleinnahmen daran tragen. In den neuen Ländern werden die Renten und Sozialleistungen diese Einnahmen überwiegen, in den alten Bundesländern spielen Kapital- und Mieteinnahmen ebenfalls eine Rolle. Ähnlich wie oben wirken sich auch die Belastungen für das Wohnen (siehe Abbildung 6) auf die Erwerbsneigung der Frauen aus. Im Westen ist der positive Anstieg mit der Miethöhe stark, auch im Osten ist solch eine Tendenz nach den ersten Mietsteigerun-

<sup>22</sup>die Variable kann z.B. durch Scheidungen oder Alimente auch negativ sein, siehe Abschnitt 3

gen zu beobachten, wengleich der Zusammenhang mit den noch deutlich niedrigeren Mietverpflichtungen wesentlich geringer ist.

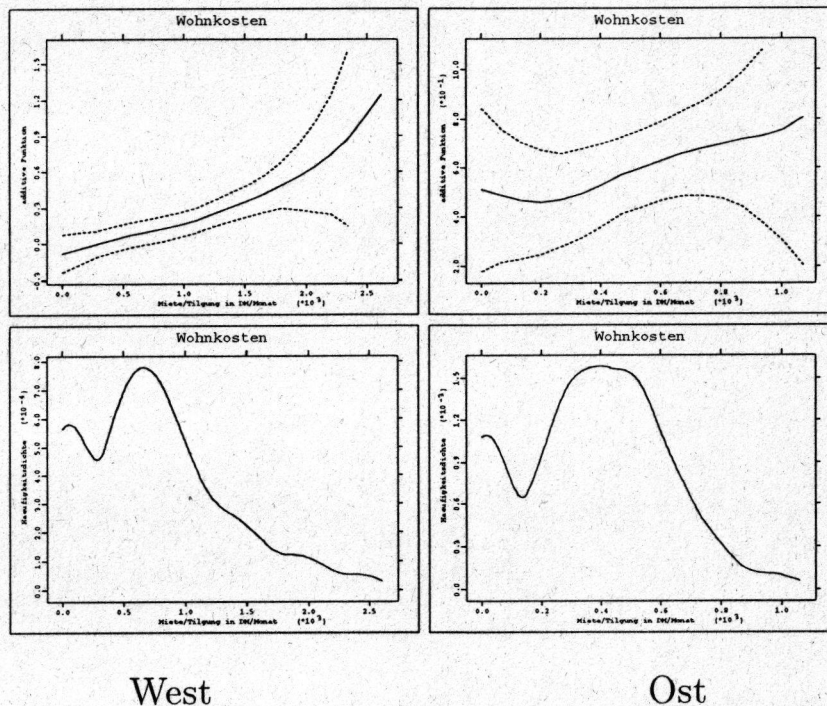
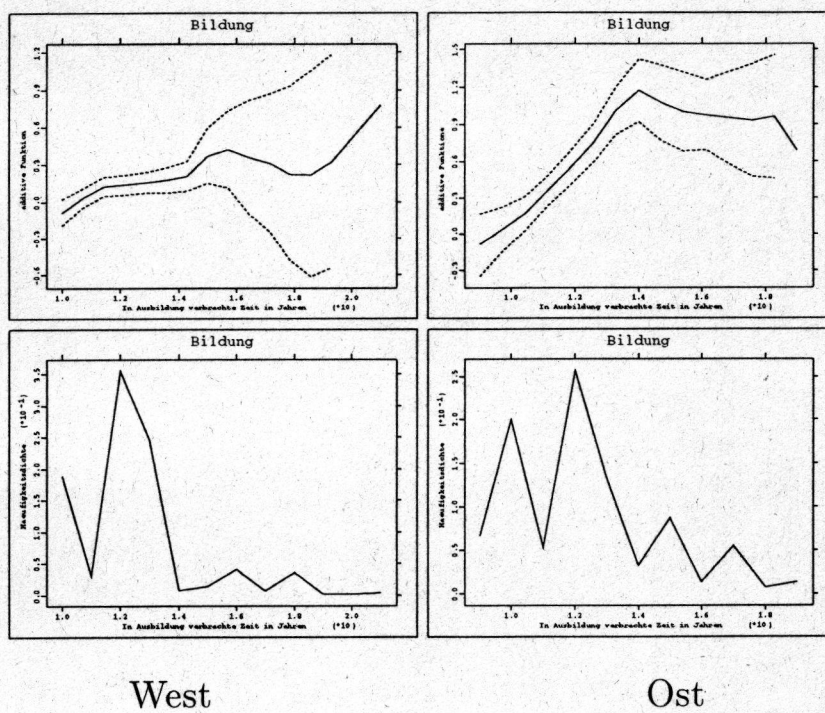
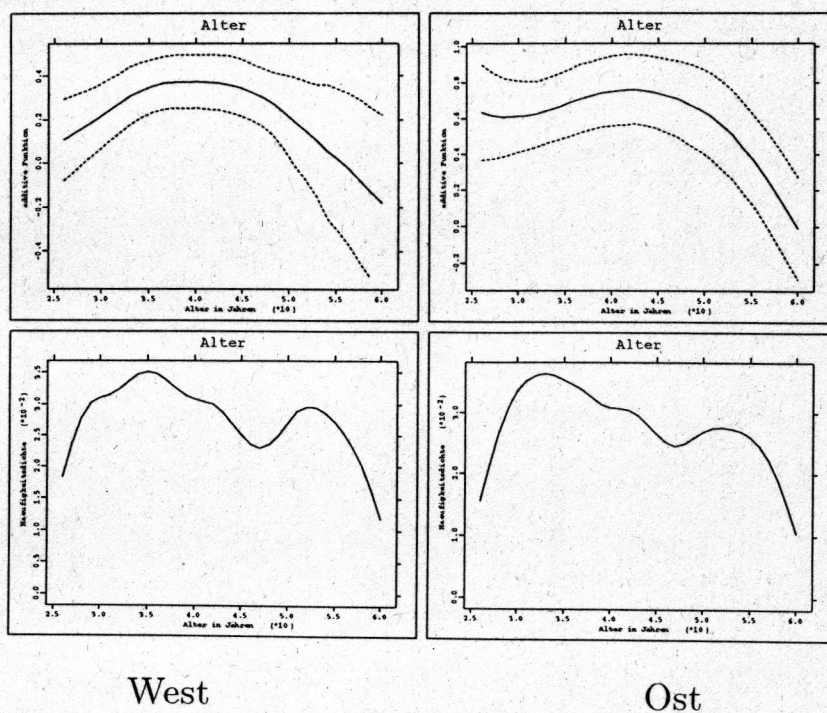


Abbildung 6: Variable *Kaltmiete/Tilgung*.

Der Einfluß der Variable Bildungsstand (Abbildung 7) auf die Berufstätigkeit ist wenig überraschend. Die Stichprobe im Westen bestätigt, daß, wer mehr in seine Ausbildung investiert hat, eher arbeiten geht. Der Verlauf der Kurve ist überwiegend ansteigend. Etwas abweichend davon ist der Anstieg unter den gut ausgebildeten Frauen in den neuen Bundesländern leicht negativ. Dies ist wie in Abbildung 2 auf die Humankapitalverluste und die Schwierigkeit, einen Arbeitsplatz für Hochqualifizierte zu finden, zurückzuführen.

In dieser Stichprobe liegt die höchste Erwerbsbeteiligung im Westen Deutschlands im Alter von etwa 35 bis 45 Jahren (siehe Abbildung 8). Bei jüngeren Frauen liegt sie etwas darunter, nach dem 45. Lebensjahr nimmt sie stark ab. Der Funktionsverlauf erscheint negativ quadratisch. Im Osten ist die Erwerbstätigkeit bis Ende 40 sehr hoch und sinkt dann rapide ab. Eine Delle etwas niedrigeren Arbeitsangebots liegt bei den etwa 30-jährigen Frauen, wo zumeist kleine Kinder vorhanden sind. Insofern werden Erkenntnisse der beschreibenden Statistik bestätigt. Gegen Ende des Erwerbslebens sinkt in ähnlichem Maße der Anteil der arbeitenden Frauen wie in den alten Ländern.

Abbildung 7: Variable *Bildungsstand*.Abbildung 8: Variable *Alter*.

Da von den Entlassungswellen zuerst die älteren Arbeitnehmerinnen betroffen waren, ist hier nichts mehr von der früher vorhandenen hohen Erwerbsquote zu sehen.

Auf eine Darstellung der Arbeitslosenquote ist verzichtet worden. Es ist keine Aussage darüber machbar, ob die hohen Erwerbsquoten auch mit hohen Arbeitslosenquoten einhergehen, viele Frauen auf dem Arbeitsmarkt also „schuld“ seien an hoher Arbeitslosigkeit. Zwar sind die Erwerbsbeteiligung wie auch die Arbeitslosigkeit im Osten deutlich höher als im Westen, aber innerhalb der beiden Gebiete ist kein Einfluß eines regionalen Konjunkturindikators auszumachen. Dies kann aber auch an der bereits hohen Aggregation der Arbeitslosenquoten auf der Ebene der Bundesländer liegen. Es sind keine genaueren Herkunftsdaten über die Personen aus dem SOEP zu erhalten, um z.B. auf einzelne Arbeitsamtsbezirke einzugehen. Leider sind auch die Daten des SOEP 1993 aus Datenschutzgründen zum letzten Mal mit der regionalen Herkunft der befragten Personen versehen gewesen.

## 7 Zusammenfassung

Mit dieser Auswertung konnten erhebliche Differenzen im Erwerbsverhalten zwischen den ost- und westdeutschen Frauen festgestellt werden. Bei den Frauen der alten Bundesländer war ein „lehrbuchmässiges“ Verhalten zu beobachten, Überraschungen blieben aus. Auffällige Unterschiede der ostdeutschen Frauen dazu bestanden insbesondere bei den Variablen Lohn, Kinder und Einkommen des Ehemanns. Diese vor allem von finanziellen Dingen bestimmten Variablen scheinen Thesen einer größeren sozialen Funktion der Erwerbstätigkeit für die ostdeutschen Frauen zu bestätigen. Ähnliche Verhältnisse in beiden Teilen Deutschlands herrschten bei Miete, Bildungsstand, Arbeitslosigkeit und sonstigem Haushaltseinkommen. Die Differenzen sind überwiegend auf die Umbruchsituation zurückzuführen. So ist z.B. die Lohnkurve eindeutig dem Transformationsprozess geschuldet. Wenn auch mittelfristig die Erwerbsbeteiligung im Osten höher bleiben wird, ist doch langfristig bei Erreichen des westdeutschen Lohnniveaus und der Arbeitslosenrate mit einer weitgehenden Angleichung zu rechnen.

Zu diesem differenzierten Bild trug wesentlich die Verwendung des semiparametrischen Regressionsmodells bei. Insbesondere bei so wichtigen Einflußfaktoren wie Lohn, Kinder, Bildung oder Einkommen des Ehemannes wären so genaue Erkenntnisse mit einem linearen Regressionsmodell wie dem Probit nicht möglich gewesen. So konnten einige überraschende Zusammenhänge erkannt werden.

### Danksagung

Ich bedanke mich bei Herrn Dipl.-Math. Stefan Sperlich, Wirtschaftswissenschaftliche Fakultät der Humboldt-Universität zu Berlin und Sonderforschungsbereich 373 „Quantifikation und Simulation ökonomischer Prozesse“, für den Beistand in Theorie und rechen technischer Umsetzung des Integrations schätzers.

### Literaturverzeichnis

- Berndt, E.R. (1990) *The practice of Econometrics: classic and contemporary*, Addison-Wesley, Reading (Mass.).
- Chen, R.; Härdle, W.; Linton, O.B.; Severance-Lossin, E. (1995) *Nonparametric Estimation of Additive Separable Regression Models*, SFB 373 Discussion Paper Nr. 50/95, Berlin.
- Deutsches Institut für Wirtschaftsforschung (1993) *Das Sozio-ökonomische Panel*, Benutzerhandbuch Bd. I und II, Version November 1993, DIW, Berlin.
- Franz, W. (1991) *Arbeitsmarktökonomik*, Springer-Verlag, Heidelberg.
- Härdle, W. (1990) *Applied Nonparametric Regression*, Econometric Monograph Series 19, Cambridge University Press.
- Härdle, W.; Hall, P. (1993) *On the Backfitting Algorithm for Additive Regression Models*, in: *Statistica Neerlandica*, 47, S. 43-57.
- Härdle, W.; Klinke, S.; Turlach, B.A. (1995) *XploRe - an interactive statistical computing environment*, Springer-Verlag, New York.
- Härdle, W.; Turlach, B.A. (1992) *Semiparametric Approaches to Dimension Reduction*, Humboldt-Universität zu Berlin, Institut für Statistik und Ökonometrie, Discussion Paper 1/92, Berlin.
- Hastie, T.J.; Tibshirani, R.J. (1991) *Generalized Additive Models*, Monographs on Statistics and Applied Probability 43, Chapman and Hall, London.
- Heckman, J.J. (1979) *Sample Selection Bias as a Specification Error*, *Econometrica*, 47, S. 153-161.
- Heckman, J.J. (1993) *What Has Been Learned About Labor Supply in the Past Twenty Years?* *American Economic Review*, 83, Papers and Proceedings, S. 116-121.



- Holst, E.; Schupp, J. (1991) *Frauenerwerbstätigkeit in den neuen und alten Bundesländern - Befunde des Sozio-ökonomischen Panels*, DIW, Discussion Paper Nr. 37, Berlin.
- Holst, E.; Schupp, J. (1994) *Erwerbsbeteiligung und Erwerbsorientierung von Frauen in West- und Ostdeutschland 1990-1993*, DIW, Discussion Paper Nr. 90, Berlin.
- Kempe, W. (1995) *Nichtparametrische Regression in Generalisierten Additiven Modellen und ihre ökonomische Anwendung*, Humboldt-Universität zu Berlin, Diplomarbeit, Berlin.
- Küsters, U.; Arminger, G. (1989) *Programmieren in GAUSS*, Gustav Fischer Verlag, Stuttgart, New York.
- Linton, O.B.; Härdle, W. (1995) *Estimation of Additive Regression Models with Links*, SFB 373 Discussion Paper Nr. 48/95, Berlin.
- Linton, O.B.; Nielsen, J.P. (1995) *A kernel method of estimating structured non-parametric regression based on marginal integration*, *Biometrika*, **82**, S. 93-100.
- McCullagh, P.; Nelder, J.A. (1989) *Generalized Linear Models*, Monographs on Statistics and Applied Probability 37, Chapman and Hall, London.
- Merz, J. (1987) *Das Arbeitsangebot verheirateter Frauen in der BRD - Eine mikroökonomische Analyse unter Berücksichtigung des selectivity bias*, In: Krupp, H.-J.; Hanefeld, U. (1987) *Lebenslagen im Wandel: Analysen 1987*, Campus, Frankfurt/Main, Mannheim.
- Merz, J. (1990) *Femal Labor Supply: Labor Force Participation, Market Wage Rate and Working Hours of Married and Unmarried Women in the Federal Republic of Germany*, *Jahrbücher für Nationalökonomie und Statistik*, **207**, S. 240-270.
- Mroz, T.A. (1987) *The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions*, *Econometrica*, **55**, S. 765-799.
- Pischner, R.; Rohwer, G. (1994) *RZOO. Ein Retrievalprogramm für das Sozio-ökonomische Panel*, Version 94.1 Wellen A-I, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin.
- Pissarides, C.A. (1990) *Equilibrium Unemployment Theory*, Basil Blackwell, Oxford.
- Ronning, G. (1991) *Mikroökonomie*, Springer-Verlag, Berlin.

- Schupp, J.; Wagner, G.G. (1995) *The German Socio-Economic Panel: a Database for Longitudinal International Comparisons*, Innovation, **8**, S. 95-108.
- Severance-Lossin, E.; Sperlich, S. (1995) *Estimation of Derivative for Additive Seperable Models*, SFB 373 Discussion Paper Nr.60/95, Berlin.
- Statistisches Jahrbuch (1992, 1993, 1994) Statistisches Bundesamt, Wiesbaden.
- Statistisches Jahrbuch der DDR (1990) Statistisches Amt der DDR, Berlin.
- Stobernack, M. (1991) *Das Arbeitsangebot in der Bundesrepublik Deutschland - Eine Schätzung auf der Basis des Sozio-ökonomischen Panels*, Jahrbücher für Nationalökonomie und Statistik, **208**, S. 625-641.
- Strøm, S.; Wagenhals, G. (1991) *Female Labour Supply in the Federal Republic of Germany*, Jahrbücher für Nationalökonomie und Statistik, **208**, S. 575-595.

