



INSTITUT FÜR STATISTIK  
SONDERFORSCHUNGSBEREICH 386



Gartner:

## Die Ersetzung fehlender Werte: Ein Test alternativer Methoden mit Makrodaten

Sonderforschungsbereich 386, Paper 216 (2000)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



# Die Ersetzung fehlender Werte: Ein Test alternativer Methoden mit Makrodaten

Hermann Gartner \*

18. Oktober 2000

## Zusammenfassung

In dieser Arbeit werden die Auswirkungen verschiedener Methoden zur Ersetzung von fehlenden Werten auf das Ergebnis einer Regressionsanalyse untersucht. Grundlage ist eine Untersuchung von Klasen (2000) über die Unterschiede im Zusammenhang zwischen Unterernährung und Kindersterblichkeit in Afrika und Süd-asien. In dem Makro-Datensatz, welcher 101 Entwicklungsländer umfasst, fällt etwa ein Drittel der 273 Beobachtungen weg, da für verschiedene verwendete Variablen die Werte fehlen. Die so verloren gegangenen Informationen sollen in diese Untersuchung genutzt werden. Hierzu werden zwei verschiedene Verfahren verwendet. Zum einen werden mit einem linearen Regressionsmodell alle Variablen wechselseitig aufeinander regressiert und mit Hilfe der so geschätzten Koeffizienten die fehlenden Werte prognostiziert. Um nicht eine zu starke Korrelation der Werte untereinander vorzutäuschen, werden die ersetzten Werte um einen zufälligen Störterm ergänzt. Zum anderen wird ein multiple imputations Verfahren verwandt, in welchem mit einem Markov-Chain-Monte-Carlo-Verfahren mehrere vervollständigte Datensätze generiert werden, welche dann weiter bearbeitet werden. Durch die Auswertung mehrerer vervollständigter Datensätze wird eine höhere Effizienz der Schätzer erreicht. Es zeigt sich, dass die Ergebnisse von Klasen (2000), etwa dass höhere Einkommen, Frauenbildung, geringere Fertilität, der Zugang zu sanitären Einrichtungen und eine geringere Bevölkerungsdichte mit geringeren Unterernährungsraten korreliert sind, robust gegenüber der Ersetzung der fehlenden Werte sind.

## 1 Einführung

In den meisten empirischen Arbeiten zu ökonomischen Fragen wird auf das Problem fehlender Werte nicht eingegangen. Wenn bei einigen Variablen Werte fehlen, wird meist die entsprechende Beobachtung aus der Untersuchung ausgeklammert. Damit gehen Informationen verloren, die effizienter genutzt werden könnten. Insbesondere bei Querschnittsdaten über mehrere Länder, wie sie auch hier verwendet werden, ist das Problem der fehlenden Werte oft gravierend. So zum Beispiel auch bei der Querschnittsuntersuchung

---

\*Ludwig-Maximilians-Universität München, Volkswirtschaftliches Institut, Seminar für Empirische Wirtschaftsforschung

von Mankiw, Romer und Weil (1992) zur Wachstumstheorie, in welcher eine Reihe von Beobachtungen (insbesondere von Entwicklungsländern) wegfallen, weil Werte fehlen.

Auf Grundlage einer Untersuchung von Klasen (2000) über den Zusammenhang von Kindersterblichkeit und Unterernährung sollen hier die aufgrund fehlender Werte verloren gegangenen Informationen genutzt werden und geprüft werden, ob sich die Ergebnisse aus der genannten Untersuchung bestätigen. Um alle Informationen im Datensatz zu nutzen, ist es erforderlich die fehlenden Werte durch Schätzungen zu ersetzen. In dieser Arbeit werden hierzu zwei verschiedene Verfahren angewandt. Zum einen ein Standardverfahren, welches eine lineare Regression anwendet, zum anderen ein moderneres Verfahren, welches mehrere vervollständigte Datensätze mittels eines Markov-Chain-Monte-Carlo-Verfahren generiert.

Die Arbeit ist wie folgt aufgebaut. Zunächst werden wichtige Ergebnisse aus der Untersuchung, worauf sich diese Arbeit bezieht, vorgestellt. Danach werden die beiden Ersetzungsverfahren diskutiert und die Ergebnisse der Anwendung auf die Untersuchung aus Klasen(2000) vorgestellt.

## 2 Das Modell mit fehlenden Werten

In Klasen (2000) wird untersucht, weshalb in Südasien relativ hohe Unterernährung zu beobachten ist, die Kindersterblichkeit jedoch vergleichsweise gering ist, in Afrika südlich der Sahara aber eine höhere Kindersterblichkeit mit einer geringeren beobachteten Unterernährung einhergeht.

Hierzu wird im Rahmen einer Regressionsanalyse versucht, die Erklärungsfaktoren für Variablen, welche Unterernährung und Kindersterblichkeit messen, zu finden. Datengrundlage sind Makrodaten von 101 Entwicklungsländern. Für viele Länder liegen Beobachtungen zu mehreren Zeitpunkten vor, so dass der Datensatz Panelcharakter hat. Insgesamt liegen somit 273 Beobachtungen vor.

Als Faktoren, welche Unterernährung und Kindersterblichkeit beeinflussen, werden die Bevölkerungsdichte, die Fertilitätsrate, die Qualifikation der Frauen, der Zugang zu sanitären und medizinischen Anlagen, die Stilldauer sowie das Sozialprodukt betrachtet. Die Bevölkerungsdichte wirkt nachteilig auf die Unterernährung und Kindersterblichkeit weil Krankheiten sich leichter ausbreiten und weil die im Land zur Verfügung stehenden Ressourcen und Nahrung knapper werden. Die Fertilitätsrate erhöht Unterernährung und Kindersterblichkeit, da bei einer höheren Fertilitätsrate die Zahl der Kindern, die eine Mutter zu versorgen hat, größer ist und somit die Mittel des Haushaltes sich auf mehr Kinder aufteilen. Eine höhere Bildung der Mutter führt dazu, dass deren Kinder besser versorgt und bei Krankheit besser gepflegt werden. Sie wirkt sich also positiv auf Unterernährung und Kindersterblichkeit aus. Die Qualifikation der Mutter wird durch die Alphabetisierungsquote der Frauen erfasst. Die Stilldauer ist von Interesse, weil eine lange Stilldauer auf eine schlechte Nahrungsversorgung hindeutet, denn wenn genügend andere Nahrung für das Kind vorhanden wäre, würde die Muttermilch viel früher abgesetzt. Es wird daher eine Variable verwendet, welche den Anteil der Kinder, die im 20 bis 23 Monat noch gestillt werden. Ein höheres Sozialprodukt erhöht die Ressourcen, über welche die Haushalte verfügen, und senkt Unterernährung und Kindersterblichkeit. Zur vertieften Diskussion der Einflussfaktoren wird auf Klasen (2000) verwiesen.

Zentrale Ergebnisse aus Klasen (2000) sind in den Tabellen 2 und 3 zu finden. Fixe Effekte der Regionen werden durch Dummyvariablen erfasst. Unterschieden werden die Regionen: Ostasien/Pazifik, Osteurope/Zentralasien, Mittlerer Osten/Nordafrika, Kari-

## 2 Das Modell mit fehlenden Werten

	N	Mittelwert	Standard- abweichung	Fehlend Anzahl	Prozent
geringes Geburtsgewicht	207	13,59	8,11	66	24,18
Bevölkerungs- dichte	268	87,30	131,86	5	1,83
ln(BIP)	258	7,44	0,76	15	5,49
Alphabeitisierung d. Frauen	247	60,03	28,68	26	9,52
Stilldauer20-23	133	41,17	20,34	140	51,28
Fertilitätsrate	268	4,77	1,68	5	1,83
Sanitärzugang	194	54,00	29,03	79	28,94
mäßiges/starkes Untergewicht	188	5,76	6,04	85	31,14
mäßiges Untergewicht	243	20,99	15,00	30	10,99
mäßige/starke Wachstumshem- mung	216	30,57	15,87	57	20,88
Impfung	205	67,99	24,21	68	24,91

Tabelle 1: Beschreibung des Datensatzes

bik, Südasien, Subsahara und Lateinamerika, wobei letztere als Referenzkategorie Verwendung findet. Um zeitliche Veränderungen aufzufangen, werden die Perioden vor 1985, 1985 bis 1989, 1990 bis 1994 und 1995 und später unterschieden.

In Tabelle 2 werden Unterernährungsmodelle geschätzt. Die Ergebnisse entsprechen den theoretischen Erwartungen. In den ersten drei Spalten ist die abhängige Variable der Anteil der untergewichtigen Neugeborenen an allen Neugeborenen. In Spalte (3) wurde ein reduziertes Modell geschätzt, bei dem die Fertilitätsrate ausgelassen wurde, da diese stark von der Bildung der Frauen abhängt. Bei der Regression in Spalte (1) fallen fast 90 Beobachtungen von den potentiell 273 Beobachtungen durch fehlende Werte weg. Durch die Hinzunahme der Stilldauer in Spalte (2) fällt über die Hälfte der Beobachtungen weg.

Die Spalten (4) und (5) zeigen das Ergebnis eines Regressionsmodells, in der der Anteil der Kinder, welche für ihr Alter sehr klein sind, zu erklären ist. Gehemmtes Wachstum ist ein Indiz für chronische Unterernährung. Bei schlechter Ernährung passt sich der Körper durch langsames Wachstum an. In Regression (5) wurden die Variablen Sanitärzugang und Fertilitätsrate weg gelassen, wodurch sich die Zahl der verwendbaren Beobachtung um 22 auf 190 erhöhte. Die letzten drei Regressionen versuchen den Anteil derer, die für ihr Alter zu wenig wiegen, zu erklären. Untergewicht tritt bei akuter Unterernährung auf. Auch hier fallen zahlreiche Beobachtungen weg, wenn die zusätzlichen Variablen Sanitärzugang und geringes Geburtsgewicht hinzugezogen werden.

Ein Ergebnis von Klasen (2000) ist, dass die Ursache von höherer Kindersterblichkeit in Afrika unter anderem in höheren Fertilitätsraten, geringerem Einkommen, geringerer Bevölkerungsdichte und schlechterer Impfschutz zu finden ist. Der Einfluss dieser Faktoren auf die Sterblichkeit der unter 5-jährigen wurde mit einer linearen Regression geschätzt, deren Ergebnisse in Tabelle 3 zu sehen sind.

Auch hier fallen durch die Hinzunahme der Regressoren geringes Geburtsgewicht und

Untergewicht gegenüber des Modells in Spalte (1) zahlreiche Beobachtungen weg. Die Hinzunahme dieser Variablen ändert nicht die Ergebnisse der Regression und deren Einfluss ist nicht signifikant. Ein Regressionsmodell, das nur die Unterernährungsvariablen enthält ist in Spalte (4) angegeben. Hier ist nur der Koeffizient Untergewicht signifikant. In Spalte (5) sind wieder die Ergebnisse eines Modells in reduzierter Form angegeben, in welchem die Fertilitätsrate ausgelassen wurde.

Tabelle 1 gibt einen Überblick über die Menge der Beobachtungen, die fehlen. Von der Variable Stilldauer fehlt über die Hälfte der Beobachtungen. Von den Unterernährungs- und Sterblichkeits-Variablen, also die zu erklären Variablen, fehlen 10-30%. Von allen Variablen, die in der Tabelle aufgelistet sind, fehlen 24 % der Werte. Dennoch fallen bei den Regressionen zum Teil über die Hälfte der Beobachtungen weg. ein großer Teil der Informationen bleibt also ungenutzt.

### 3 Regressionsverfahren

Ein einfaches Verfahren um die fehlenden Werte zu ersetzen ist die Anwendung eines linearen Regressionsmodelles. Dieses Verfahren ist im Statistikpaket SPSS implementiert. Hierbei werden jene Beobachtungen, bei denen keine der Variablen fehlt, verwendet um alle Variablen wechselseitig aufeinander zu regressieren.

Bei der Prognose der fehlenden Werte ist zu beachten, dass keine zu starke Korrelation vorgetäuscht wird. Dies wäre der Fall, wenn die fehlenden Werte durch jene Werte ersetzt werden, die auf der Regressionsgerade liegen, also durch  $\mathbf{X}\beta$ . Es ist daher ein Störterm  $\mathbf{u}$  zu ergänzen, welcher aus einer normalverteilten Zufallsverteilung gezogen wird. Die Varianz des Störtermes wird mit Hilfe des vollständigen Datensatzes geschätzt.

Zur Schätzung der fehlenden Werte wird der Panel-Charakter des Datensatzes genutzt. Gibt es für ein Land zwei oder mehr Beobachtungen aus verschiedenen Jahren und fehlt beispielsweise in einem Jahr der Wert des Sozialproduktes, liegt der wahre Wert in der Nähe der Werte, die beobachtet wurden. Um diese Informationen zu nutzen, wurde ein Modell mit fixen Effekten für die Länder, für die wenigstens zwei Beobachtungen vorlagen, verwendet. Die Länder, für die nur eine Beobachtung vorliegt, wurden zu Regionen, wie sie in den Unterernährungs- und Sterblichkeits-Modellen verwendet wurden, zusammengefasst. Eine Reihe von Variablen sind Anteilsgrößen. Sie können also nur zwischen null und eins liegen. Prognostizierte Werte, die über diese Grenzen hinaus gingen, wurden auf null bzw. eins gesetzt, um plausible Werte sicherzustellen. Dies traf zu auf die Variablen Impfung, geringes Geburtsgewicht, starkes Untergewicht, Alphabetisierung der Frauen, Stilldauer20-23 zu.

Die Tabellen 4 und 5 zeigen die Ergebnisse der Regressionsmodelle, die auch in den Tabellen 2 und 3 geschätzt wurden, nun jedoch mit dem vervollständigten Datensatz. Die Dummyvariable für die Region Osteuropa/Zentralasien, die bei der Verwendung des unvollständigen Datensatzes bei drei Regressionen wegen Multikollinearität wegfallen musste, kann bei dem vervollständigten Datensatz ohne Probleme in den Regressionen verwendet werden.

Die Koeffizienten von Karibik und Mittlerer Osten/Nordafrika wechseln bei Klassen(2000) in manchen Modellvarianten das Vorzeichen. In den Regressionsmodellen aus der Tabelle 4 bleibt das Vorzeichen immer negativ. Ebenso ist das Vorzeichen von Alphabetisierung der Frauen immer negativ. Es hat, anders als bei den Regressionen mit dem unvollständigen Datensatz, immer das erwartete Vorzeichen. Eine höhere Qualifikation der Frauen verringert den Anteil der unterernährten Kinder.

## 4 Multiple Imputation

Auch bei den Sterblichkeitsmodellen finden Vorzeichenwechsel statt. Während in Tabelle 3 der Koeffizient Osteuropa/Zentralasien in den Spalten (2) und (3) positive Vorzeichen aufweist, in den anderen Spalten ein negatives, haben in der Schätzung mit dem vervollständigten Datensatz in Tabelle 5 die Osteuropa/Zentralasien-Koeffizienten alle ein negatives Vorzeichen. Das un plausible negative Vorzeichen vor geringes Geburtsgewicht in der 4. Spalte der Schätzung aus Klasen(2000) verschwindet bei der Schätzung mit dem vervollständigten Datensatz.

Ansonsten bleiben die Größenordnungen der Koeffizienten im Wesentlichen erhalten. Die Ergebnisse aus Klasen (2000) können also bestätigt werden. Ein überraschendes Ergebnis ist, dass das Bestimmtheitsmaß bei den Regressionen mit allen Beobachtungen sinkt. Eine mögliche Erklärung ist, dass gerade bei den Ländern, deren Beobachtungen häufig fehlen, die Messfehler der beobachteten Werte größer sind, und dadurch die Anpassung sich verschlechtert hat.

## 4 Multiple Imputation

Ein Verfahren, das von Joseph L. Schafer propagiert wird, ist das Multiple-Imputation-Verfahren<sup>1</sup>. Mit Hilfe eines Markov-Ketten-Monte-Carlo-Prozesses werden mehrere vollständige Datensätze erzeugt, die getrennt voneinander analysiert werden. Im zweiten Schritt werden die Ergebnisse dann verknüpft.

Vorausgesetzt wird in diesem Verfahren, dass die Variablen gemeinsam normalverteilt sind und dass die Werte zufällig fehlen (*missing at random*). Damit ist gemeint, dass die Tatsache, ob ein bestimmter Wert fehlt oder nicht, nur von den beobachteten Werten und von den Parametern des datengenerierenden Prozesses abhängt, nicht aber von den fehlenden Werten selber.

Um die Bedingung formal zu verdeutlichen sind einige Größen zu definieren. Der Datensatz  $Y$  bestehen aus den beobachteten Werten  $Y_{obs}$  und aus den fehlenden Werten  $Y_{mis}$ , so dass  $Y = (Y_{obs}, Y_{mis})$ . An welcher Position Werte fehlen, wird durch die Matrix  $R$  angezeigt. Sie hat den Eintrag 1, wenn der entsprechend Wert in  $Y$  fehlt, ansonsten den Wert 0. Der unbeobachtbare Parametervektor des datengenerierenden Prozesses ist  $\xi$ . Die *missing at random*-Bedingung lautet damit:

$$P(R|Y_{obs}, Y_{mis}, \xi) = P(R|Y_{obs}, \xi) \quad (1)$$

Das Ersetzungsverfahren ist eine iterative Prozedur, die zunächst einen Startwert für die Parameter der zugrundegelegten Verteilung erfordert, die bei Schafer (1997) mit einem Erwartungs-Maximierungs-Verfahren (EM) geschätzt werden. Das Verfahren, um einen vervollständigten Datensatz zu erzeugen, verläuft dann wie folgt. Zunächst werden zufällige Werte für die fehlenden Daten, gegeben die Startwerte der Parameter, aus der angenommenen Verteilung gezogen.

$$Y_{mis}^{(t+1)} \sim P(Y_{mis}|Y_{obs}, \theta^{(t)}) \quad (2)$$

Dann werden neue Parameter aus einer bayesianischen Verteilung auf Basis der beobachteten *und* der geschätzten Daten gezogen.

---

<sup>1</sup>Zur Diskussion und genaueren Beschreibung dieses Verfahrens wird auf Schafer (1997, 1998) verwiesen

$$\theta^{(t+1)} \sim P(\theta | Y_{mis}^{(t+1)}, Y_{obs}, \theta^{(t)}) \quad (3)$$

Mit den neuen Parameterschätzungen wird der Prozess wiederholt. Es entsteht so eine Markov-Kette, welche sich nach einer Reihe von Iterationen stabilisiert. Nach einer genügend großen Zahl von Iterationen werden die fehlenden Werte  $Y_{mis}$  aus der Verteilung gezogen. Diese bilden mit  $Y_{obs}$  den ersten vervollständigten Datensatz. Dieses Verfahren wird mehrfach wiederholt, so dass am Ende mehrere vervollständigte Datensätze vorliegen. Diese Datensätze werden dann getrennt analysiert und deren Ergebnisse nach den Regeln, wie sie in Rubin (1987) beschrieben sind, zusammengeführt.

Die Generierung der Datensätze lässt sich mit dem unter Windows lauffähigen Programm NORM durchführen, welches von der Homepage von J.L. Schafer geladen werden kann (unter <http://www.stat.psu.edu/~jls/>).

Zur Generierung der fehlenden Werte konnte hier nicht auf einzelne Länderdummies wie im Regressionsverfahren zurückgegriffen werden. Der hier verwendete Ersetzungsalgorithmus ist weniger robust gegenüber Kollinearitäten in den Variablen, die bei der Verwendung der zahlreichen Länderdummies auftritt.

Um die fehlenden Werte des hier diskutierten Datensatzes zu ersetzen wurden 1000 Iterationen durchgeführt. Um zu guten Ergebnissen zu kommen genügen bereits wenige Datensätze. So wurden vier Datensätze generiert, auf welche wieder die Regressionen aus den Tabellen 2 und 3 durchgeführt wurden. Die Kombination der Ergebnisse ist in den Tabellen 6 und 7 dargestellt.

Auch hier ändern sich in den Unterernährungsmodellen (Tabelle 6) wieder die Vorzeichen der Koeffizienten Karibik und Mittlerer Osten/Nordafrika und Alphabetisierung der Frauen. Allerdings sind nur beim Koeffizienten Karibik die Vorzeichen in allen Modellvarianten negativ. Bei den anderen beiden Koeffizienten bleibt in je einer Variante ein positives Vorzeichen erhalten.

Bei den Sterblichkeitsmodellen (Tabelle 7) ändern sich hier nur die Vorzeichen des Dummies für die Jahre 90-94. Allerdings ist deren Standardabweichung auch sehr hoch. Auffällig ist hier, dass sich die Standardabweichung der Regionaldummies stark erhöht. Die Größenordnungen und die Vorzeichen der anderen Koeffizienten bleiben in beiden bei der Verwendung dieses Datensatzes im Wesentlichen erhalten. Auch mit dem Multiple-Imputation-Verfahren können also die Ergebnisse von Klasen (2000) gestützt werden.

## 5 Ausblick

Es hat sich gezeigt, dass durch die Ersetzung der fehlenden Werte aus der Untersuchung von Klasen (2000) dessen Ergebnisse erhärtet werden konnten. Darüberhinaus führte die Verwendung der vervollständigten Datensätze zu Ergebnissen, bei denen unplausible Vorzeichen verschwanden.

Durch die Ersetzung der fehlenden Werte können noch eine Reihe von Variablen zugänglich gemacht werden, die in den bisherigen Untersuchungen nicht verwendet wurden, da dadurch auf noch mehr Beobachtungen hätte verzichtet werden müssen. So sind in den Datensätzen unter anderem noch Informationen über die Einkommensverteilung, den Zugang zu Wasser und Gesundheitseinrichtungen enthalten, deren Einfluss auf Unterernährung und Kindersterblichkeit noch einer weiteren Untersuchung bedarf.

## Literatur

**Klasen, S.** (2000). Malnourished and surviving in South Asia, better nourished and dying in Africa: What can explain this puzzle? SFB 386, Ludwig-Maximilians-Universität München, Discussion Paper 214.

**Mankiw, G. N., Romer, D. und Weil, D.** (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economic*, (May), 407–437.

**Rubin, D.** (1987). *Multiple Imputation for Nonresponse Surveys*. J. Wilson & Sons, New York.

**Schafer, J. L.** (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall, London.

**Schafer, J. L.** (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. Techn. Ber., Pennsylvania State University.

## 6 Anhang



Abhängige Variable	1 geringes Ge- burtsgewicht	2 geringes Ge- burtsgewicht	3 geringes Ge- burtsgewicht	4 mäki- ge/starke Wachstums- hemmung	5 mäki- ge/starke Wachstums- hemmung	6 starkes Untergewicht	7 mäki- ges/starkes Untergewicht	8 mäki- ges/starkes Untergewicht
Ostasien/Pazifik	2,500 (1,03)	2,753 (1,39)	2,331 (1,03)	6,411 (2,58)	6,031 (2,59)	2,702 (0,82)	15,650 (1,73)	15,580 (1,75)
Osteurope/Zentralasien	-3,817 (2,51)		-4,389 (2,46)		-8,633 (3,47)	0,320 (1,55)		-2,905 (2,80)
Mittlerer Osten/Nordafrika	-2,029 (1,26)	1,211 (1,55)	-2,158 (1,26)	0,432 (2,88)	-3,533 (2,72)	-1,409 (1,06)	0,164 (2,28)	-3,083 (2,15)
Karibik	-2,023 (1,75)	0,874 (2,66)	-2,251 (1,74)	-8,821 (3,56)	-11,601 (2,95)	-0,360 (1,08)	-1,294 (3,99)	-1,549 (2,35)
Südasien	15,304 (1,89)	13,305 (2,04)	15,001 (1,89)	9,683 (3,99)	10,620 (3,74)	10,103 (1,44)	20,642 (4,13)	20,677 (2,96)
Subsahara	-2,866 (1,14)	-2,150 (1,39)	-2,800 (1,14)	-7,936 (2,57)	-5,079 (2,31)	-0,337 (0,88)	5,946 (2,17)	2,957 (1,81)
d9094	0,088 (0,86)	-0,411 (1,00)	0,148 (0,86)	-3,737 (2,02)	-1,724 (1,86)	-0,347 (0,69)	0,563 (1,53)	-0,204 (1,44)
d8589	1,616 (0,96)	2,150 (1,11)	1,837 (0,95)	-3,850 (2,26)	-2,199 (2,09)	-0,759 (0,79)	-1,136 (1,76)	-0,620 (1,65)
pre1985	1,978 (1,26)	4,037 (3,99)	2,154 (1,25)	-1,686 (2,39)	2,765 (2,15)	0,520 (0,82)	-0,190 (2,33)	0,382 (1,66)
Bevölkerungsdichte	0,019 (0,00)	0,018 (0,00)	0,018 (0,00)	0,023 (0,01)	0,017 (0,01)	0,007 (0,00)	0,022 (0,01)	0,021 (0,00)
Fertilitätsrate	0,354 (0,38)	0,180 (0,52)		3,259 (0,84)			1,704 (0,66)	
Alphabetisierung d. Frauen	-0,045 (0,02)	0,005 (0,03)	-0,055 (0,02)	0,026 (0,05)	-0,121 (0,04)	-0,066 (0,01)	-0,030 (0,04)	-0,150 (0,03)
ln(BIP)	-3,136 (0,68)	-3,302 (1,20)	-3,338 (0,61)	-7,602 (1,74)	-12,191 (1,31)	-2,722 (0,46)	-2,876 (1,49)	-6,960 (0,97)
Stillddauer20-23		0,068 (0,03)						
Sanitärzugang				-0,140 (0,04)			-0,090 (0,04)	
geringes Geburtsgewicht							0,209 (0,15)	
Konstante	35,429 (6,50)	31,378 (11,64)	39,231 (4,56)	78,426 (15,60)	128,598 (9,57)	28,639 (3,43)	31,172 (13,88)	76,564 (7,19)
Angepasstes $R^2$	0,773	0,760	0,770	0,725	0,702	0,746	0,828	0,776
Ramsey Test P:	0,000	0,000	0,000	0,831	0,591	0,000	0,402	0,321
N	186	114	187	138	190	170	136	217

Tabelle 2: Unterernährungsmodelle mit unvollständigen Datensatz

## 6 Anhang

	1	2	3	4	5
Abhängige Variable	Sterblichkeit der unter 5jährigen				
Ostasien/Pazifik	-15,504 (8,64)	-10,236 (8,97)	-15,656 (11,32)		-18,300 (8,72)
Osteuropa/Zentralasien	-0,619 (13,09)	4,399 (20,22)	6,471 (20,23)		-11,610 (12,70)
Mittlerer Osten/Nordafrika	-24,417 (10,10)	-13,807 (10,65)	-14,714 (10,93)		-25,911 (10,24)
Karibik	-23,053 (12,19)	-20,066 (14,23)	-17,403 (14,31)		-26,772 (12,31)
Südasien	-18,682 (13,98)	-17,309 (18,17)	-28,026 (21,46)		-25,106 (14,01)
Subsahara	4,605 (9,19)	19,747 (9,86)	20,833 (10,01)		7,747 (9,27)
d9094	1,799 (6,59)	-0,922 (6,80)	-2,326 (6,92)		3,708 (6,66)
d8589	12,263 (7,46)	1,453 (7,86)	2,082 (8,12)		15,377 (7,48)
pre1985	16,007 (10,68)	36,488 (24,30)	35,071 (24,58)		23,424 (10,52)
Bevölkerungsdichte	0,011 (0,02)	-0,030 (0,03)	-0,043 (0,03)		-0,001 (0,02)
Fertilitätsrate	8,669 (3,06)	6,549 (3,24)	5,863 (3,39)		
Alphabetisierung d. Frauen	-0,923 (0,17)	-0,733 (0,18)	-0,767 (0,19)		-1,154 (0,15)
ln(BIP)	-34,238 (5,38)	-28,516 (6,14)	-25,713 (6,61)		-40,617 (4,94)
Impfung		-0,300 (0,16)	-0,295 (0,17)		
geringes Geburtsgewicht		1,361 (0,65)	1,373 (0,67)	-0,249 (0,77)	
mäßiges Untergewicht			0,341 (0,42)	2,665 (0,42)	
Konstante	374,991 (50,76)	334,246 (61,52)	314,667 (66,88)	48,936 (8,54)	475,826 (36,53)
Angepasstes $R^2$	0,786	0,811	0,811	0,306	0,779
Ramsey Test P:	0,014	0,362	0,448	0,000	0,077
N	198	167	156	176	199

Tabelle 3: Sterblichkeitsmodelle mit unvollständigen Datensatz

*Für alle Tabellen gilt: Ausgelassene Kategorien sind Lateinamerika und 1995+; Standardabweichungen in Klammern*

Abhängige Variable	1 geringes Ge- burtsgewicht	2 geringes Ge- burtsgewicht	3 geringes Ge- burtsgewicht	4 mäki- ge/starke Wachstums- hemmung	5 mäki- ge/starke Wachstums- hemmung	6 starkes Untergewicht	7 mäki- ges/starkes Untergewicht	8 mäki- ges/starkes Untergewicht
Ostasien/Pazifik	3,094 (1,05)	2,294 (1,06)	2,957 (1,06)	11,509 (1,98)	11,120 (2,06)	3,072 (0,71)	14,940 (1,76)	16,021 (1,93)
Osteurope/Zentralasien	0,059 (1,63)	0,903 (1,62)	-1,082 (1,60)	-5,452 (3,09)	-9,166 (3,10)	1,281 (1,07)	-0,991 (2,70)	-5,051 (2,90)
Mittlerer Osten/Nordafrika	-1,882 (1,17)	-1,692 (1,15)	-1,385 (1,17)	-5,269 (2,27)	-3,940 (2,27)	-1,082 (0,79)	-1,317 (1,98)	-0,837 (2,13)
Karibik	-3,224 (1,53)	-2,117 (1,54)	-3,477 (1,55)	-11,639 (2,88)	-12,593 (3,00)	-0,806 (1,04)	-0,220 (2,54)	-2,860 (2,81)
Südasien	9,764 (1,74)	9,011 (1,72)	9,467 (1,76)	10,848 (3,35)	10,331 (3,41)	8,347 (1,18)	12,600 (3,04)	16,864 (3,19)
Subsahara	-2,717 (1,08)	-2,751 (1,06)	-2,065 (1,07)	-7,135 (2,04)	-4,967 (2,08)	-0,091 (0,72)	0,141 (1,80)	1,225 (1,94)
d9094	0,406 (0,89)	0,470 (0,87)	0,680 (0,89)	-1,013 (1,68)	-0,187 (1,73)	-0,459 (0,60)	-2,068 (1,47)	-0,980 (1,62)
d8589	1,669 (1,00)	2,037 (0,99)	2,114 (1,00)	-3,203 (1,88)	-1,669 (1,94)	-0,944 (0,67)	-4,221 (1,65)	-1,705 (1,81)
pre1985	0,808 (1,01)	0,570 (0,99)	1,650 (0,97)	1,479 (1,91)	4,358 (1,89)	0,376 (0,65)	-3,382 (1,67)	0,211 (1,77)
Bevölkerungsdichte	0,022 (0,00)	0,021 (0,00)	0,021 (0,00)	0,020 (0,01)	0,015 (0,01)	0,008 (0,00)	0,018 (0,01)	0,024 (0,01)
Fertilitätsrate	0,929 (0,35)	0,961 (0,34)	0,961 (0,34)	3,202 (0,65)	3,202 (0,65)	3,120 (0,58)	3,120 (0,58)	3,120 (0,58)
Alphabetisierung d. Frauen	-0,026 (0,02)	-0,019 (0,02)	-0,046 (0,02)	-0,047 (0,04)	-0,126 (0,03)	-0,059 (0,01)	-0,030 (0,03)	-0,133 (0,03)
ln(BIP)	-2,216 (0,60)	-1,814 (0,60)	-2,917 (0,55)	-7,228 (1,17)	-9,902 (1,06)	-2,416 (0,37)	-2,295 (1,03)	-6,407 (0,99)
Stilddauer20-23	0,604 (0,00)	0,619 (0,00)	0,594 (0,00)	0,653 (0,091)	0,622 (0,099)	0,667 (0,000)	0,719 (0,968)	0,649 (0,127)
Sanitärzugang				-0,023 (0,03)			-0,031 (0,03)	
geringes Geburtsgewicht							0,483 (0,11)	
Konstante	24,460 (5,61)	18,810 (5,74)	34,804 (4,12)	74,243 (10,65)	111,343 (8,00)	26,064 (2,76)	18,810 (9,56)	72,319 (7,48)
Angepasstes $R^2$								
Ramsey Test P:								
N	273	273	273	273	273	273	273	273

Tabelle 4: Unternährungsmodelle, fehlende Werte mit Regressionsverfahren ersetzt

## 6 Anhang

Abhängige Variable	1	2	3	4	5
Sterblichkeit der unter 5jährigen					
Ostasien/Pazifik	-15,210 (8,87)	-17,849 (9,02)	-23,792 (10,09)		-17,070 (9,16)
Osteurope/Zentralasien	-10,264 (13,79)	-10,610 (13,61)	-10,380 (13,59)		-25,791 (13,76)
Mittlerer Osten/Nordafrika	-14,627 (9,89)	-10,116 (9,94)	-9,097 (9,96)		-7,830 (10,10)
Karibik	-16,054 (12,91)	-11,514 (12,85)	-11,441 (12,84)		-19,496 (13,32)
Südasien	-2,147 (14,69)	-9,749 (16,17)	-14,210 (16,51)		-6,165 (15,16)
Subsahara	4,101 (9,15)	8,968 (9,18)	9,068 (9,17)		12,981 (9,22)
d9094	-1,152 (7,50)	-1,641 (7,41)	-0,690 (7,43)		2,584 (7,70)
d8589	8,145 (8,44)	5,159 (8,39)	6,848 (8,48)		14,210 (8,60)
pre1985	12,543 (8,54)	11,684 (8,44)	13,162 (8,50)		23,996 (8,39)
Bevölkerungsdichte	0,013 (0,02)	-0,016 (0,03)	-0,023 (0,03)		-0,004 (0,02)
Fertilitätsrate	12,634 (2,92)	10,590 (2,98)	9,180 (3,17)		
Alphabetisierung d. Frauen	-0,384 (0,15)	-0,335 (0,15)	-0,317 (0,15)		-0,655 (0,14)
ln(BIP)	-33,031 (5,06)	-30,290 (5,13)	-29,296 (5,18)		-42,585 (4,70)
Impfung		-0,101 (0,09)	-0,117 (0,09)		
geringes Geburtsgewicht		1,156 (0,56)	0,899 (0,59)	0,322 (0,67)	
mäßiges Untergewicht			0,416 (0,32)	2,198 (0,34)	
Konstante	315,143 (47,31)	295,229 (49,95)	289,447 (50,07)	59,376 (7,40)	455,938 (35,54)
Angepasstes $R^2$	0,662	0,670	0,671	0,239	0,638
Ramsey Test P:	0,211	0,045	0,069	0,000	0,103
N	273	273	273	273	273

Tabelle 5: Sterblichkeitsmodelle, fehlende Werte mit Regressionsverfahren ersetzt

Abhängige Variable	1 geringes Ge- burtsgewicht	2 geringes Ge- burtsgewicht	3 geringes Ge- burtsgewicht	4 mäki- ge/starke Wachstums- hemmung	5 mäki- ge/starke Wachstums- hemmung	6 starkes Untergewicht	7 mäki- ges/starkes Untergewicht	8 mäki- ges/starkes Untergewicht
Ostasien/Pazifik	2,988 (1,00)	2,928 (0,98)	2,872 (1,00)	8,798 (2,38)	7,974 (2,51)	2,609 (0,75)	16,116 (1,79)	16,069 (1,86)
Osteurope/Zentralasien	-2,749 (1,45)	-2,473 (1,53)	-3,355 (1,41)	-2,311 (3,03)	-7,339 (3,17)	-0,313 (1,07)	1,582 (2,87)	-2,581 (2,71)
Mittlerer Osten/Nordafrika	-2,471 (1,19)	-2,226 (1,25)	-2,388 (1,20)	-0,903 (2,29)	-3,017 (2,54)	-1,298 (1,13)	0,569 (2,28)	-0,703 (2,38)
Karibik	-2,144 (1,74)	-1,775 (1,87)	-2,310 (1,74)	-10,327 (2,88)	-11,456 (2,93)	-0,770 (1,08)	-0,806 (2,49)	-2,185 (2,59)
Südasien	14,656 (3,13)	14,183 (3,65)	14,292 (3,12)	9,458 (3,30)	8,581 (3,46)	9,650 (1,45)	15,414 (3,62)	16,925 (3,50)
Subsahara	-2,979 (1,44)	-3,010 (1,40)	-2,715 (1,48)	-6,561 (2,03)	-6,261 (2,09)	-0,028 (0,88)	1,268 (1,82)	1,326 (1,93)
d9094	0,262 (0,81)	0,286 (0,81)	0,417 (0,80)	-1,645 (1,63)	-1,339 (1,70)	-0,493 (0,72)	-1,709 (1,42)	-1,153 (1,48)
d8589	1,443 (0,93)	1,513 (0,95)	1,665 (0,93)	-3,210 (1,83)	-2,131 (1,87)	-1,115 (0,68)	-3,389 (1,62)	-2,065 (1,70)
pre1985	0,689 (1,08)	1,030 (1,18)	1,097 (1,08)	0,820 (2,15)	3,112 (1,89)	0,610 (0,74)	-1,362 (1,61)	0,793 (1,63)
Bevölkerungsdichte	0,019 (0,00)	0,019 (0,00)	0,019 (0,00)	0,024 (0,01)	0,018 (0,01)	0,007 (0,00)	0,023 (0,01)	0,024 (0,01)
Fertilitätsrate	0,550 (0,32)	0,639 (0,34)	0,639 (0,34)	3,347 (0,94)	3,347 (0,94)	2,728 (0,66)	2,728 (0,66)	2,728 (0,66)
Alphabetisierung d. Frauen	-0,047 (0,02)	-0,042 (0,02)	-0,062 (0,02)	0,026 (0,04)	-0,132 (0,03)	-0,057 (0,02)	-0,014 (0,04)	-0,134 (0,03)
ln(BIP)	-2,495 (0,72)	-2,154 (0,83)	-2,892 (0,65)	-8,363 (1,56)	-12,951 (1,14)	-2,739 (0,42)	-4,563 (1,25)	-8,190 (1,03)
Stilddauer20-23	0,022 (0,03)	0,022 (0,03)	0,022 (0,03)	-0,132 (0,03)	-0,132 (0,03)	-0,064 (0,03)	-0,064 (0,03)	-0,064 (0,03)
Sanitärzugang								
geringes Geburtsgewicht								
Konstante	29,845 (6,65)	25,779 (8,07)	36,116 (5,13)	82,012 (15,46)	134,550 (8,82)	28,618 (3,17)	40,000 (11,59)	85,547 (7,64)
Angepasstes $R^2$								
Ramsey Test P:								
N	273	273	273	273	273	273	273	273

Tabelle 6: Unterernährungsmodelle, fehlende Werte mit Multiple-Imputation-Verfahren ersetzt

## 6 Anhang

Abhängige Variable	1	2	3	4	5
Sterblichkeit der unter 5jährigen					
Ostasien/Pazifik	-12,398 (7,90)	-15,128 (7,74)	-17,507 (10,20)		-14,317 (7,93)
Osteurope/Zentralasien	-0,993 (11,92)	0,724 (11,91)	0,554 (11,77)		-10,951 (11,39)
Mittlerer Osten/Nordafrika	-26,683 (121,28)	-15,855 (116,51)	-16,273 (110,18)		-25,389 (131,26)
Karibik	-26,024 (140,81)	-27,557 (131,94)	-27,441 (136,12)		-28,851 (143,08)
Südasi en	-15,004 (166,98)	-26,233 (214,00)	-29,180 (302,72)		-21,061 (168,54)
Subsahara	0,143 (88,46)	6,542 (71,75)	6,293 (72,14)		4,197 (88,42)
d9094	1,131 (6,15)	1,217 (5,99)	1,458 (6,05)		3,583 (6,26)
d8589	10,475 (6,97)	4,260 (7,03)	4,759 (7,48)		14,124 (7,18)
pre1985	9,739 (12,41)	2,471 (12,79)	2,902 (11,91)		16,457 (11,39)
Bevölkerungsdichte	-0,001 (0,02)	-0,024 (0,02)	-0,026 (0,03)		-0,010 (0,02)
Fertilitätsrate	8,963 (3,11)	6,289 (3,13)	5,972 (2,97)		
Alphabetsierung d. Frauen	-0,912 (0,16)	-0,732 (0,16)	-0,733 (0,16)		-1,159 (0,13)
ln(BIP)	-35,953 (4,59)	-33,633 (4,57)	-32,888 (5,26)		-42,523 (4,64)
Impfung		-0,339 (0,14)	-0,327 (0,15)		
geringes Geburtsgewicht		1,183 (0,51)	1,149 (0,51)	-0,210 (0,66)	
mäßiges Untergewicht			0,136 (0,49)	2,840 (0,37)	
Konstante	388,928 (44,04)	382,406 (46,47)	375,739 (56,87)	51,865 (7,53)	491,960 (34,66)
N	273	273	273	273	273

Tabelle 7: Sterblichkeitsmodelle, fehlende Werte mit Multiple-Imputation-Verfahren ersetzt