

Simulation Extrapolation Seminarbericht

Andreas Hölzl
Betreuung Eva Endres

9. Januar 2015

Inhaltsverzeichnis

1 Einführung zu SIMEX	2
1.1 Erklärung zu SIMEX	2
1.2 Simex-Algorithmus	11
1.3 Theorie zu SIMEX	12
1.4 Simulationsstudien zu Simex	12
2 Bestimmung des Messfehlers	14
2.1 Kalibrierung	14
2.2 Messwiederholung	14
2.3 Modellbasiert	15
3 SIMEX bei semiparametrischer Regression	17
3.1 Splines	17
3.1.1 Wahl der Knoten	21
3.1.2 Problematik des Glättungsparameters	21
3.2 Anwendung auf reale Daten	22

Kapitel 1

Einführung zu SIMEX

1.1 Erklärung zu SIMEX

Im Folgenden wird versucht, die Idee hinter SIMEX zuerst intuitiv zu erklären, bevor das ganze Vorgehen im nächsten Kapitel formalisiert wird. Betrachten wir ein lineares Regressionsmodell

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon$$

wobei ϵ normalverteilt sei. Die Simulation eines derartigen Modells ist in der Grafik [1.1](#) visualisiert. Wie man sehen kann, sind die theoretische Regressionsgerade und die aus den Daten gefundene Regressionsgerade sehr ähnlich.

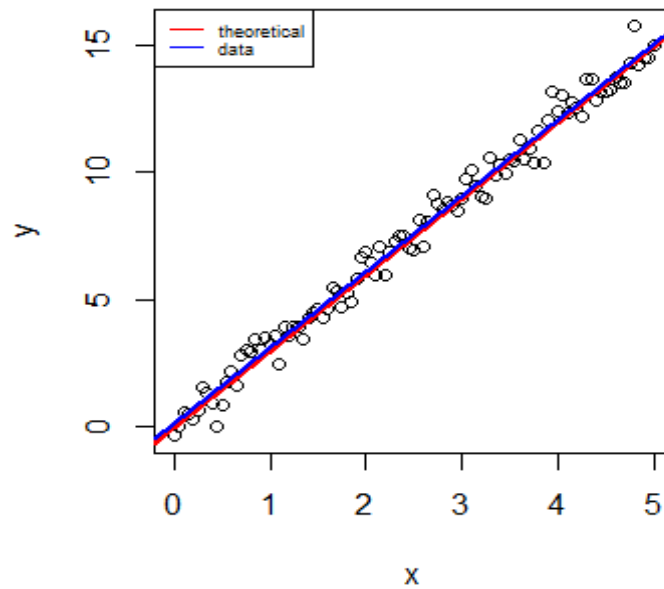


Abbildung 1.1: Schätzung der Regressionsgleichung bei Fehlern in der abhängigen Variable

Bei Vorliegen eines additiven Messfehlers wird auch die unabhängige Variable nicht genau gemessen. Unter der Annahme einer Normalverteilung für den Messfehler und einer festen Standardabweichung für diesen kann dieser Messfehler dann simuliert werden, was in [Abbildung 1.2](#) abgebildet ist. Im Weiteren werde davon ausgegangen, dass die Messfehlerstandardabweichung σ_x dieser Normalverteilung bekannt sei. In einem weiteren Kapitel wird darauf eingegangen, wann und wie diese Messfehlerstandardabweichungen bestimmt werden können.

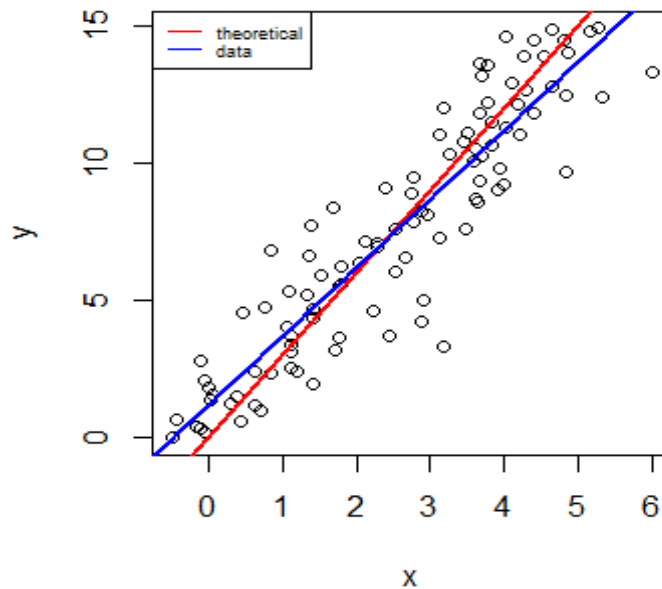


Abbildung 1.2: Schätzung der Regressionsgleichung bei Fehlern in der unabhängigen Variable

Das Vorliegen des Messfehlers in der unabhängigen Variable führt zu einem Bias der Regressionsschätzung, was in vorherigen Vorträgen auch schon theoretisch hergeleitet wurde. Es gilt dabei

$$\beta_{est} = \frac{\beta_{true}}{1 + \frac{\sigma}{\sigma_{x^2}}}$$

der Parameter β_1 wird also systematisch unterschätzt. Eine Fehlerkorrektur im linearen Modell wäre also direkt analytisch machbar, allerdings wollen wir hier nur das lineare Modell als Beispiel nehmen um ein Verfahren herzuleiten, dass auch für viel allgemeinere Modelle angewendet werden kann.

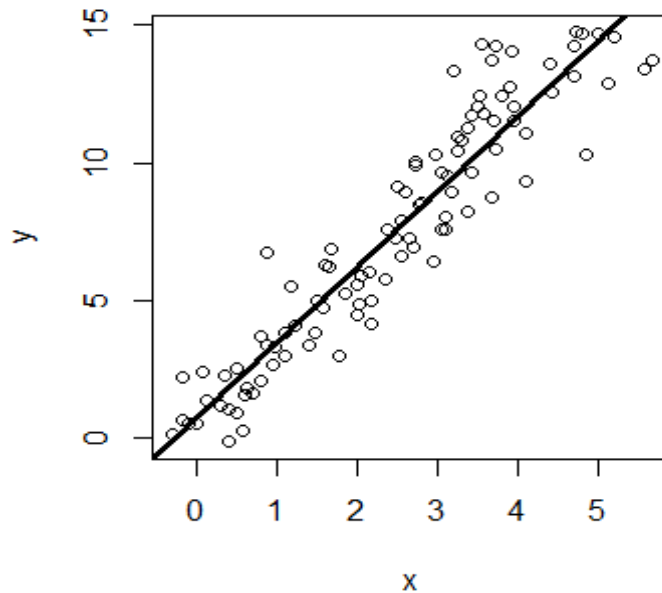


Abbildung 1.3: Naiver Schätzer mit Messfehlern

Wir wollen im Folgenden betrachten, wie sich das Modell verändert, wenn man einen Messfehler zu der unabhängigen Variable hinzufügt. Hierfür fügen wir zu den bereits mit Messfehlern gemessenen Daten einen weiteren aus der gleichen Verteilung wie der ursprüngliche Messfehler stammenden Messfehler hinzu. Die Standardabweichung dieses neuen Messfehlers entspricht dabei $\lambda_1 * \sigma_x$, wobei λ_1 zunächst frei wählbar ist, hier beispielsweise gleich 1.

Da der Messfehler $\lambda_1 * \sigma_x$ zu den bereits mit Messfehler σ_x gemessenen Daten hinzugefügt wurde, gilt für den gesamten Messfehler der so erzeugten Daten $X^*(\lambda_1) : \sigma_{x^*(\lambda_1)}^2 = \sigma^2(1 + \lambda_1)$.

In der Grafik 1.4 sind in schwarz die mit Messfehler gemessenden Punkte und die dazu gefittete Regressionsgerade abgebildet und in blau die Punkte, die entstehen, wenn ein weiterer Messfehler $\lambda_1 * \sigma_x$ zu diesen Punkten

hinzugefügt wird. Die blaue Regressionsgerade zeigt das dazu gefittete Regressionsmodell an.

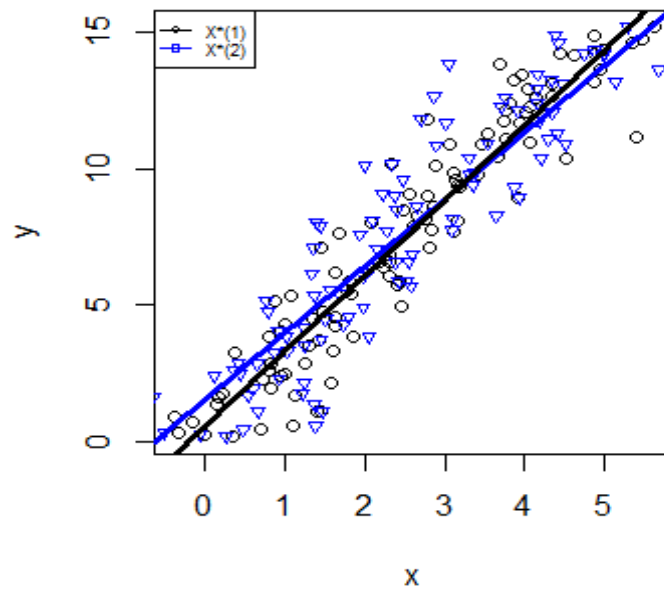


Abbildung 1.4: Hinzufügen eines zusätzlichen Messfehlers

Dies wird noch einmal gemacht, dieses Mal aber mit einer anderen Standardabweichung $\lambda_2 = 2$ und es wird wieder beobachtet, wie sich das lineare Modell verändert. In Grafik 1.4 sind die roten Punkte die Punkte, die entstehen, wenn der Messfehler $\lambda_2 * \sigma_x$ zusätzlich hinzugefügt wird.

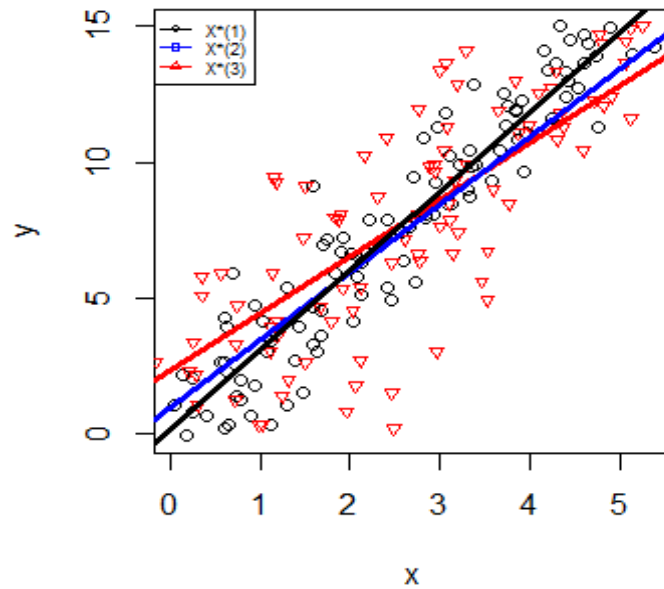


Abbildung 1.5: Hinzufügen eines weiteren zusätzlichen Messfehlers

Visuell wird schnell klar, dass der Intercept β_0 immer größer wird, während die Steigung β_1 immer kleiner wird. Man kann sich für alle dabei herauskommenden Parameter in einer Tabelle eintragen. Hier ergeben sich beispielsweise die folgenden Werte:

Parameter/ λ	0	1	2
Intercept	0.9976851	1.969627	3.077275
Steigung	2.6683550	2.394560	1.859049

Man betrachtet die sich ergebenden Werte dann in einem Plot, um zu betrachten, welcher Zusammenhang zwischen λ und den Parametern vorliegen könnte. In 1.6 ist der Plot für den Parameter β_1 aufgeführt.

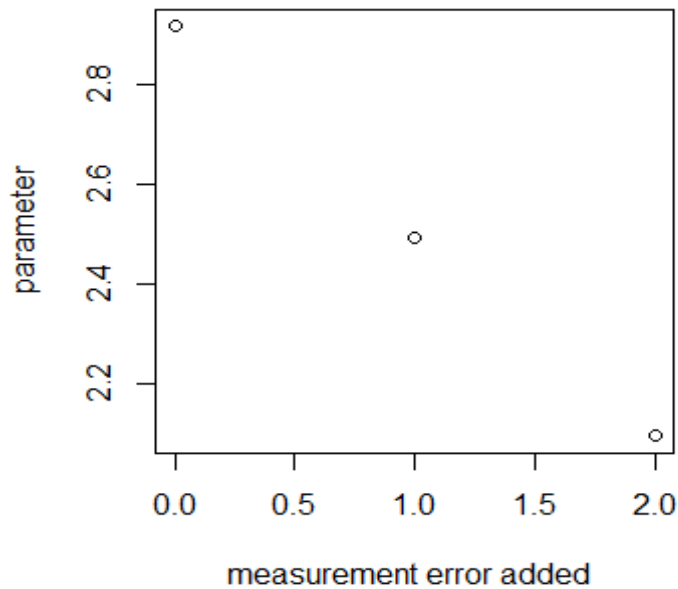


Abbildung 1.6: Parameterschätzer in Abhängigkeit von λ

Es scheint, dass hier ein lineares Modell gefittet werden kann, und genau das machen wir nun auch, um den Parameterschätzer für $\lambda = -1$ abzuschätzen, was dem Vorliegen von gar keinem Messfehler entspricht. Der Parameterschätzer für $\lambda = -1$ entspricht dabei laut unserer Interpretation von λ , dass ein der Messfehler $-\sigma_x$ zum ursprünglichen Messfehler σ_x hinzugefügt wird, was dem Messfehler 0, also dem Fehlen eines Messfehlers entspricht.

Genau dies ist die Idee hinter Simex. Man kann Messfehler nur hinzufügen und nicht wieder wegnehmen, aber man kann sich anschauen, wie das Hinzufügen von Messfehlern die Parameter verändert und über diesen Zusammenhang dann ein Modell erstellen, das dann vorhersagen kann, was passieren würde, wenn kein Messfehler vorliegen würde.

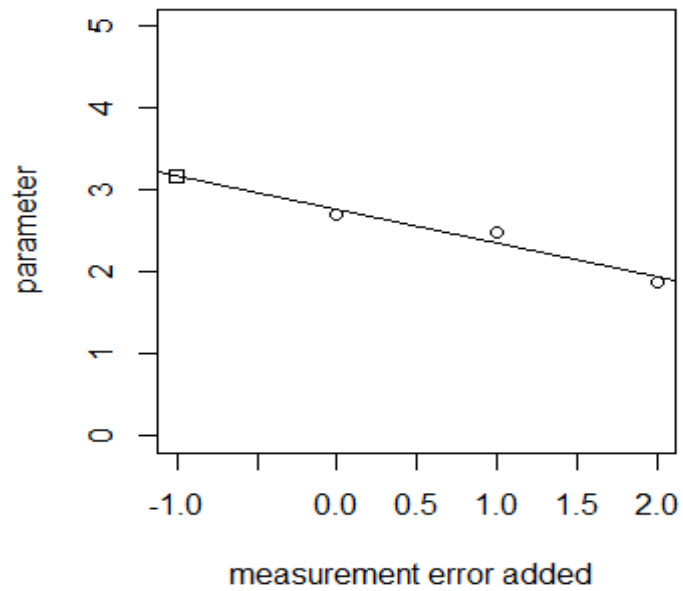


Abbildung 1.7: Schätzung des Parameters bei Vorliegen von keinem Messfehler

In der `simex`-Funktion des R-Packages `simex` wird das Ganze dann auch ähnlich visualisiert:

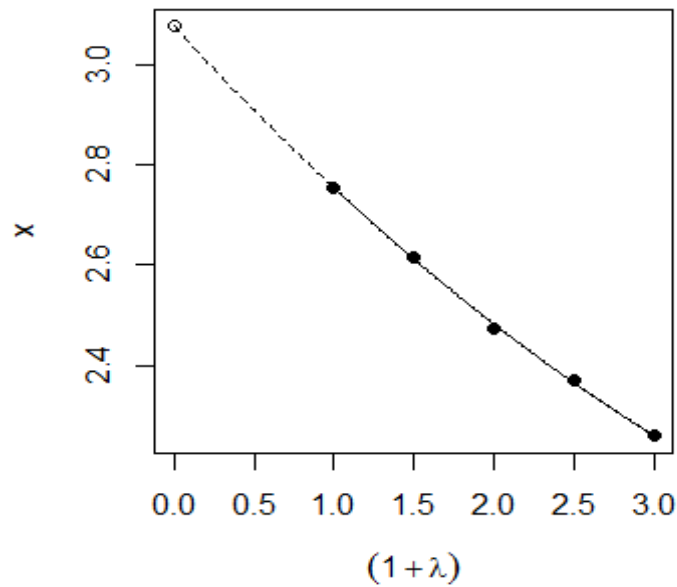


Abbildung 1.8: Schätzung des Parameters bei Vorliegen von keinem Messfehler im Simex-package

In diesem Kapitel wurde also die Idee für Simex intuitiv an einem kleinen Beispiel erläutert. Es gibt offensichtlich noch Verbesserungsschläge. So ist die Schätzung der Parameter eines Modells mit dazuaddiertem Messfehler $\lambda\sigma_x$ mit Varianz verbunden, weil der gefundene Parameter verschieden ist, je nachdem wie die Ausprägung der Messfehler tatsächlich gezogen wurde. Dieser Effekt kann aber dadurch verkleinert werden, dass man den Parameter B mal schätzt und dann mittelt.

In nächsten Abschnitt werde nun der Simex-Algorithmus noch einmal formal aufgeschrieben.

1.2 Simex-Algorithmus

Simulationsschritt

1. Simuliere Pseudo-daten $X_{b,i}(\lambda) = X_i^* + \sqrt{\lambda\sigma^2}U_{b,i}$ mit $U_{b,i} \sim N(0, 1)$
2. Mache dies B mal
3. Berechne Mittelwertschätzer für alle λ : $\beta(\lambda) = \frac{1}{B} \sum_0^B \beta_{naive}(Y_i, X_{b,i}(\lambda))$

Explorationsschritt

1. Finde ein Modell, das den Zusammenhang zwischen $(\lambda_k, \beta_k(\lambda_k))$ modelliert
2. Mache die Vorhersage dieses Modells für $\beta(-1)$

Häufige Modelle sind dabei das lineare Modell und das quadratische Modell. Es gibt hier aber keine klaren Vorschriften, man sollte nur ein Modell finden, das die Daten möglichst gut beschreibt. Angenommen dieser Wert wurde richtig vorhergesagt, so gilt es als der Regressionsparameter für diese $x_{i,b}$, für die λ gleich -1 ist und somit keine Messfehler vorliegen.

1.3 Theorie zu SIMEX

Simex wurde zunächst von Cook und Stefanski in (Cook&Stefanski, 1994) vorgestellt, zunächst ohne formalen Beweis, warum die Methode funktioniert. In (Carroll&Küchenhoff et al. 1996) wird dann aber der Beweis erbracht, dass die Simex Methode unter sehr allgemeinen Bedingungen, nämlich dass Schätzgleichungen vorliegen, den messfehlerfreien Parameter asymptotisch richtig schätzt. Diese Bedingung bedeutet, dass es zu jedem möglichen hinzuaddierten Fehler mit der gleichen Gleichung die Parameter bestimmt werden können. Im linearen Modell ist das beispielsweise die Schätzgleichung

$$\beta = (X^T X)^{-1} X^T y$$

Somit können auch viele andere Modelle, beispielsweise auch generalisierte lineare Modelle mit Simex behandelt werden.

1.4 Simulationsstudien zu Simex

Auch praktisch wurde SIMEX in verschiedenen Modellen mit Hilfe von Simulationsstudien getestet. In (Shang, 2012) wurde Simex beispielsweise zur Messfehlerbehebung in der Quantilsregression verwendet und der Autor kommt zum Schluss "A simulation study is conducted to demonstrate the performance of the SIMEX method in reducing bias and mean squared error in quantile regression with a mismeasured predictor and "the SIMEX method corrects such biases and closely reproduces conditional distributions of current test scores given past true scores". Zu einem ähnlichen Schluss kommt der Artikel (Hu&Liang, 2012): "We present theoretical justification of the estimation procedure along with a simulation study to demonstrate finite sample performance".

Aber Simex funktioniert nicht für alle Modelle. So kommen Bo Hu, Liang Li et al in (Mallick&Fung, 2002) zum Schluss "SSIMEX, on the other hand, failed to adequately adjust for the effects of random measurement error in the Cox model, even in the presence of a moderate degree of measurement error."

Simex hat also unter allgemeinen Bedingungen theoretisch bewiesene asymptotische Eigenschaften und es wurde auch in Simulationsstudien gezeigt, dass es für eine Vielzahl von Modellen gültig ist. Trotzdem wird es im Vergleich dazu, wie oft es sinnvoll wäre, es zu verwenden, relativ selten verwen-

det. Der Hauptgrund liegt darin, dass man für die Simex-Methode zuerst den Messfehler σ_x wissen muss. Diesen haben wir bisher als bekannt angenommen, er ist es aber im Normalfalls nicht. Das nächste Kapitel beschäftigt sich also damit, wie σ_x bestimmt werden kann.

Kapitel 2

Bestimmung des Messfehlers

Hier werden drei Möglichkeiten aufgeführt, wie der Messfehler bestimmt werden kann.

2.1 Kalibrierung

Wenn der wahre Wert von bestimmten Messungen bekannt ist, kann dies dazu verwendet werden, um die Messfehlerverteilung zu schätzen. Die Messung zu diesem Messwert wird dann mehrfach gemacht und die Verteilung der Messpunkte dann parametrisch geschätzt, beispielsweise durch eine Normalverteilung mit bestimmter Standardabweichung. Ein Beispiel wäre ein Gerät, das den pH-Wert misst. Man weiß, dass steriles Wasser bei 25 Grad einen pH-Wert von 7 hat. Man misst also steriles Wasser bei 25 Grad mehrfach mit dem Testgerät und kann dann die Verteilung schätzen, aus der die Messfehler hervorkommen.

2.2 Messwiederholung

Ist kein wahrer Wert gegeben, aber es werden mehrere Messungen gemacht, so kann unter der Annahme, dass kein Bias vorliegt, auch die Verteilung der Messung bestimmt werden.

Für Simex gibt es in (Devanarayana&Stefanski, 2002) eine Methode, mit der im linearen Modell mehreren Messungen direkt für die Simex-Abschätzung, ohne dass vorher direkt die Parameter der Verteilung direkt bestimmt werden. Man hat also nicht nur einen mit Messfehlern behafteten Messwert X^*_i zu einer Messung, sondern mehrere mit gleichen Messfehler behaftete Messungen X^*_{ij} .

Es sollen neue Pseudodaten generiert werden, für die, damit Simex ausgeführt werden kann, gilt :

$$\text{Var}(X_{b,i}) = \sigma_x^2 * (1 + \lambda) \quad (2.1)$$

Dies soll gemacht werden, in dem Linearkombinationen der gegebenen Messwiederholungen X_{*ij} gebildet werden.

$$X_{*b,i}(\lambda) = \sum_{j=1}^{m_i} A_{bi} X_{*ij}$$

m_i beschreibt dabei die Anzahl der Messwiederholungen, die für Messung i vorliegen.

Wenn dabei für A_{ij} dabei noch die zwei Bedingungen erfüllt sind

$$\sum_1^m c_i = 1$$

und

$$\sum_1^m c_i^2 = \frac{1 + \lambda}{m}$$

so gilt 2.1 und Pseudodaten mit zusätzlichem Messfehler λ können mit diesem Verfahren erzeugt werden.

Der Beweis findet sich in (Devanarayana&Stefanski, 2002) und dort ist auch aufgeführt, wie die A_{ij} computational effizient mit Normalverteilungen simuliert werden können.

2.3 Modellbasiert

Für bestimmte Modelle ist auch implizit ein Messfehler mitgeschätzt. Dies ist beispielsweise in der Faktoranalyse der Fall. Am Beispiel einer psychologischen Untersuchung soll die einfaktorielle Faktoranalyse hier kurz erklärt werden. Es soll in einer psychologischen Untersuchung untersucht werden, welchen Einfluss die Angst vor Mathematik auf die Note hat. Die Emotion Angst kann aber nicht direkt gemessen werden, sondern nur indirekt über Fragebögen. Jeder der n Testpersonen füllt also m Fragen über diese Emotion aus. Diese seien so gestellt, dass sie alle Fragen über Angst und mit möglichst wenig Fremdeinfluss von anderen Gegebenheiten sind. Jede Frage wird auf einer Skale von 1 (stimme gar nicht zu) bis 5 (stimme sehr zu) beantwortet. Für das Modell der einfaktoriellen Faktoranalyse gilt dann

$$x_{ij} = mu_j + L * f_i + u_{ij}$$

mu_j ist dabei der Mittelwert, die bei der Frage j über alle Personen beantwortet wurde. f_i ist der wahre Wert der latenten Variable Angst und L gibt an, wie sehr die echte Variable Angst die Antwort auf die entsprechende Frage im Fragebogen beeinflusst. u_{ij} sind die übrigbleibenden Residuen.

Dieses Modell ist die einfaktorielle Faktoranalyse und eine Möglichkeit die L und f_i zu bestimmen wäre ein Maximum-Likelihood Ansatz, bei dem bspw. für die u_{ij} eine Normalverteilung angenommen wird. Bestimmt man diese Parameter nun mithilfe der Parameter, so entspricht

$$x_{ij} = x_{ij}^{theor} + u_{ij}$$

mit

$$x_{ij}^{theor} = mu_j + L * f_i$$

einem Messfehlermodell. Um den Effekt dieses Messfehlers zu bereinigen kann SIMEX angewendet werden. Die u_{ij} werden durch die eindeutige Schätzung der anderen Parameter auch eindeutig bestimmt und somit kann der Messfehler bestimmt werden. Derartige Messmodelle sind in der Psychometrie sehr verbreitet und gehören dort zur klassischen Testtheorie (KTT).

Kapitel 3

SIMEX bei semiparametrischer Regression

3.1 Splines

Splines beschreiben Funktionen der Art

$$f(x) = \sum_1^n \gamma_i B(X)$$

Grundlage der semiparametrischen Regression ist eine Basifunktion $B(X)$ mit einem bestimmten Knotenpunkt. Zu jedem der Werte X kann nun ein Wert berechnet werden, wie groß die Ausprägung dieses Punktes an der Basisfunktion mit dem entsprechenden Knotenpunkt ist. Die Designmatrix $Z = B_j(x_i)$ beschreibt dann die Ausprägung jedes Punkte x_i and der Basisfunktion mit Knoten j und ist Grundlage der Regression. Das Ergebnis der Regression ist dann ein Parameter pro Knotenpunkt und um die geschätzte Regressionsgleichung zu betrachten, wird jede Basisfunktion mit diesem Faktor multipliziert und dann alle Basisfunktionen zusammenaddiert. In 3.1 ist dies visualisiert.

Für die Schätzung von γ geht oft auch ein Glättungsparameter λ ein, der verhindern soll, dass die Schätzung sich zu sehr auf die Daten konzentriert und weniger den allgemeinen Trend findet als die speziellen Ausprägungen dieser Daten.

Es soll dann

$$(y - Z\gamma)(y - Z\gamma) + \lambda\gamma K\gamma$$

minimiert werden.

Dies geht bei Normalverteilungsannahme der Residuen über

$$\gamma = (Z^T Z + \lambda K)^{-1} Z^T y$$

Es wird also jeder einzelne Regressionsparameter bestimmt. Für jede Z und y kann für einen festen Glättungsparameter λ eindeutig der Gewichtsvektor γ gefunden.

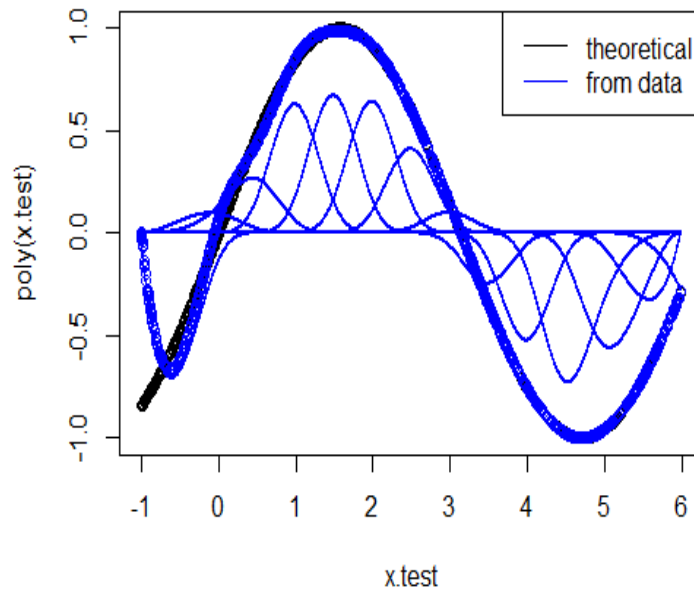


Abbildung 3.1: Visualisierung von Splines

Wir betrachten hier allerdings mit Messfehler gemessene Daten und diese könnten beispielsweise so aussehen wie in Grafik 3.2 visualisiert.

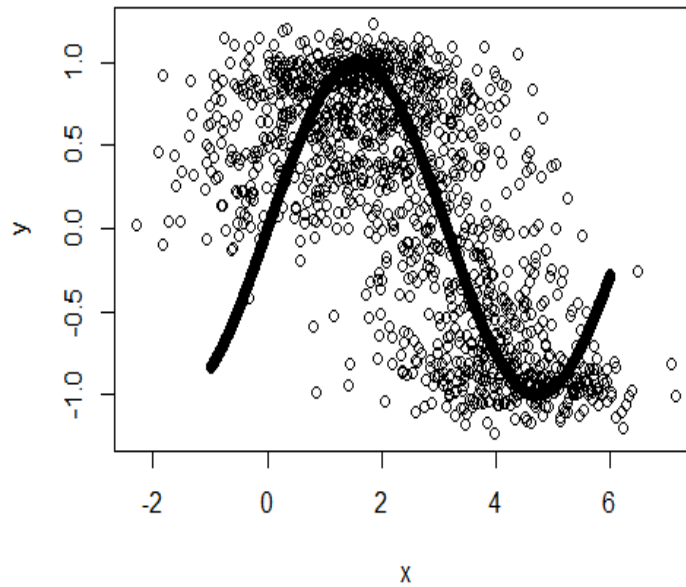


Abbildung 3.2: Hinzufügen eines zusätzlichen Messfehlers

Dabei sieht die geschätzte semiparametrische Funktion beispielsweise folgendermaßen aus:

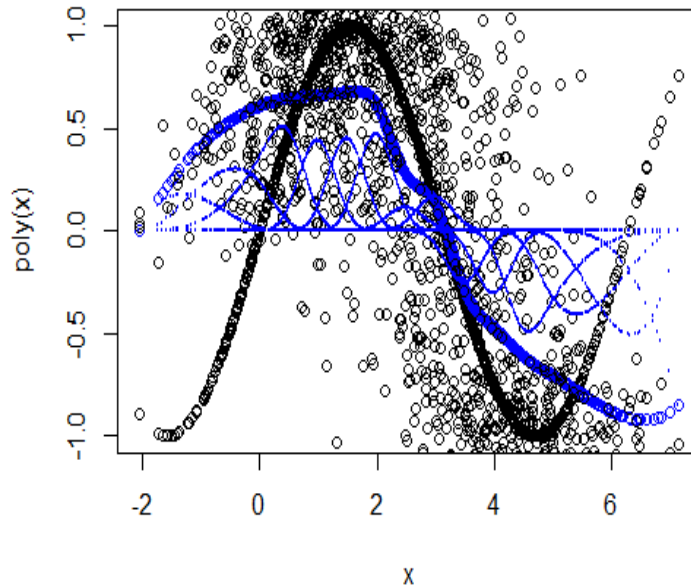


Abbildung 3.3: Splines bei Messfehlern

Wie beobachtet werden kann, wird die wahre Funktion nicht ganz richtig geschätzt, die geschätzte semiparametrische Funktion ist zu flach. Dies versuchen wir nun mit Simex zu korrigieren.

Die Idee für Simex lautet dabei:

1. Simulationsschritt : Für verschiedene λ füge aus der angenommenen Messfehlerverteilung mit der entsprechenden Standardabweichung Messfehler zur x-Variable hinzu und berechne die Designmatrix Z . Berechne jedes mal den Parametervektor γ mit

$$\gamma = (Z^T Z + \lambda K)^{-1} Z^T y$$

und middle für jedes einzelne λ die γ über die B Ziehungen.

2. Extrapolationsschritt: Stelle eine Regressionsschätzung für jeden einzelnen Spline auf und extrapoliere damit zu $\lambda = -1$ zurück. Die dabei

herauskommenden Werte sind dann die von Simex geschätzten Werte.

Es gibt allerdings noch zwei Probleme, die bei normalem Simex nicht auftreten:

3.1.1 Wahl der Knoten

Die Knoten müssen, damit Simex funktioniert, auf alle Fälle immer gleich gewählt sein. Deshalb wird es aktuell so gemacht, dass die Knoten bei den Daten der naiven Regression mit Messfehlern äquidistant gewählt werden.

3.1.2 Problematik des Glättungsparameters

Es gibt allerdings noch das Problem der Wahl des Glättungsparameters. In der Literatur wird dies für normale P-Splines zumeist durch Kreuzvalidierung gemacht. Das Problem bei Simex ist allerdings, dass für verschiedene Daten, also für die verschiedenen λ mit den verschiedenen hinzugefügten Messfehlern, verschiedene Glättungsparameter geschätzt werden würden.

Wenn man den Glättungsparametern allerdings jedes mal neu wählt, gelten die in (Carroll&Küchenhoff et al. 1996) gezeigten asymptotischen Eigenschaften nicht mehr. Es gilt dann nämlich nicht mehr die gleiche Schätzgleichung für die verschiedenen Daten, die sich ergeben, wenn man verschiedene Messfehler hinzufügt. In (Carroll&Ruppert, 2004) wurde beide Möglichkeiten, entweder den Glättungsparameter einmal auf den gegebenen Daten zu berechnen und dann immer zu verwenden oder den Glättungsparameter bei jedem Simulationsschritt neu zu berechnen, beschrieben, allerdings nicht miteinander verglichen. In (Carroll&Maca, 1999) wurde eine andere Methode zur Schätzung des Glättungsparameters für Simex vorgeschlagen, die darauf beruht, die den MSE für verschiedene Glättungsparameter abzuschätzen.

Zwischen den möglichen Verfahren - Bestimmung des Glättungsparameters einmal fest über Kreuzvalidierung, Glättungsparameter, in jedem Schritt neu bestimmen und Glättungsparameter über das Verfahren in (Carroll&Maca, 1999) abzuschätzen, gab es allerdings noch keine vergleichende Simulationsstudie und für keines dieser Verfahren gibt es Theorie, die asymptotische Eigenschaften beweisen würde. In dieser Arbeit wurde der Glättungsparameter jedes mal neu über Kreuzvalidierung bestimmt.

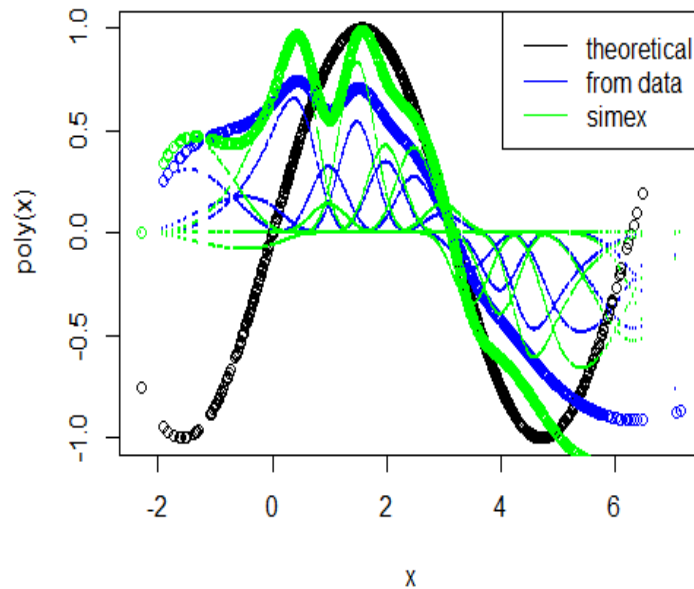


Abbildung 3.4: Splines bei Messfehlern

In 3.4 kann nun das Ergebnis der Anwendung der Simex-methode beobachtet werden. Die gefundene semiparametrische Funktion ist näher an der theoretisch angenommenen Funktion, Man muss aber trotzdem vorsichtig sein, da wir in diesem Fall keine aus Theorie hervorgehenden asymptotischen Eigenschaften der Simex-methode haben.

3.2 Anwendung auf reale Daten

Nun soll die Simex-methode für semiparametrische Daten auch angewendet werden. Hierfür werden Daten verwendet, die den Zusammenhang zwischen Angst vor Mathematik und den Noten in Mathematik untersuchen. Da Emotionen nicht direkt gemessen werden können, werden Items erhoben, auf jedes Item kann von 1 (stimme gar nicht zu) bis 5 (stimme sehr zu) geantwortet werden. Wie in der Psychologie üblich werden diese ordinalen Variablen dann als metrisch aufgefasst und eine Faktoranalyse durchgeführt,

die eine latente Variable ergibt. Zusätzlich erhalten wir in der Faktorenanalyse Information über den Messfehler, von der die tatsächlichen Messungen durchgeführt werden. Somit kann also Simex durchgeführt werden und es wird der nichtparametrische Simex durchgeführt.

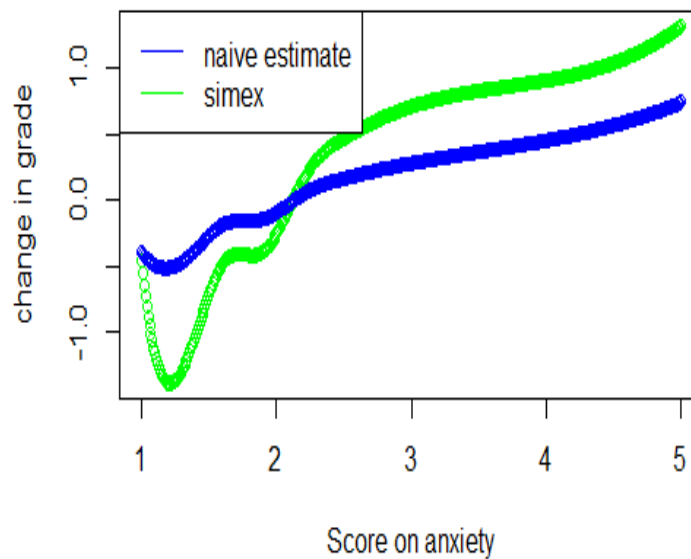


Abbildung 3.5: Anwendung der Simex-methode für Splines an echten Daten

In Grafik 3.5 werden die Ergebnisse dargestellt. Es scheint auf den ersten Blick so zu sein, dass die sonst übliche lineare Modellierung nicht ganz angebracht ist, weil die Steigung zwischen 1 und 2 anders ist als im restlichen Wertebereich. Bei der Interpretation der Ergebnisse sollte man dennoch noch vorsichtig sein, da die vorgestellte Methode noch keine bewiesenen asymptotischen Eigenschaften besitzt.

Literatur

- [Cook&Stefanski, 1994] Cook, J.R. and Stefanski, L.A. (1994) Simulation-extrapolation estimation in parametric measurement error models. *Journal of American Statistical Association*, 89, 1314 - 1328
- [Lederer&Küchenhoff,2006] Lederer, W. and Küchenhoff, H. (2006) A short introduction to the SIMEX and MCSIMEX. *RNews*, 6(4), 26 - 31
- [Carroll&Küchenhoff et al. 1996] Carroll R.,Küchenhoff, H. et al. Asymptotics for the SIMEX estimator in Nonlinear Measurement Error Model, *Journal of the American Statistical Association*; Mar 1996; 91, 433; *ABI/INFORM Global* pg. 242
- [Shang, 2012] Shang Y. (2012), Measurement Error Adjustment Using the SIMEX Method: An Application to Student Growth Percentiles, *Journal of Educational Measurement* Winter 2012, Vol. 49, No. 4, pp. 446-465
- [Hu&Liang, 2012] Hu B, Liang L. (2012), Nonparametric Multi-state Representations of Survival and Longitudinal Data with Measurement Error, *Stat Med.* 2012 September 20; 31(21)
- [Mallick&Fung, 2002] Mallick R, Fung K (2012) Adjusting for measurement error in the Cox proportional hazards regression model., *J Cancer Epidemiol Prev.* 2002;7(4):155-64
- [Devanarayana&Stefanski, 2002] Devanarayana V.; Stefanski L.(2002) Empirical simulation extrapolation for measurement error models with replicate measurements, *Statistics and Probability Letters* 59 (2002) 219-225
- [Carroll&Ruppert, 2004] Carroll R.; Ruppert D.; Crainiceanu C..(2004), *Nonlinear and Nonparametric Regression and Instrumental Variables*,

Journal of the American Statistical Association September 2004, Vol. 99, No. 467

[Carroll&Maca, 1999] Carroll R.; Maca C.; Ruppert D. (1999), Nonparametric regression in the presence of measurement error , *Biometrika* 86,3 pp. 541 - 554