

Zum Problem der Verteilungen mit Shortfall bei der Nutzung des Indikators 'Durchschnittliche Körpergröße'

Baten, Jörg; Heintel, Markus

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Baten, J., & Heintel, M. (1995). Zum Problem der Verteilungen mit Shortfall bei der Nutzung des Indikators 'Durchschnittliche Körpergröße'. *Historical Social Research*, 20(4), 135-154. <https://doi.org/10.12759/hsr.20.1995.4.135-154>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

METHODS: SOFTWARE, REVIEWS, NOTICES

Zum Problem der Verteilungen mit Shortfall bei der Nutzung des Indikators »Durchschnittliche Körpergröße«

*Jörg Baten und Markus Heintel**

Abstract: Research in economic history frequently uses human heights as a proxy for net nutrition. This anthropometric method enables historians to measure time trends and regional differences in nutritional status. However, the most widely used data sources for historical height measurements cannot be regarded as random samples of their underlying populations. In personnel records of volunteer armies, the lower side of the otherwise normal distribution is eroded by a phenomenon called »shortfall«. Because recruiting practices favoured especially tall soldiers, shorter individuals are underrepresented below a certain threshold (»truncation point«). This article explains and compares different methods of estimating the true mean and standard deviation of the underlying population from these biased data sets. We conclude that an estimator called TPE/RSMLE controlled by the K&K method has better statistical features than the frequently used QBE. As an example, the height trend in Bavaria during the late 18th century is estimated using both the TPE/RSMLE and K&K method.

* Address all communications to Jörg Baten or Markus Heintel, Ludwig-Maximilians-Universität, Volkswirtschaftliches Institut, Seminar für Wirtschaftsgeschichte, Ludwigstr. 33/TV, D-80539 München. This paper is part of an ongoing series of papers on anthropometric research in this journal, which will be continued. Markus Heintel will send the QBE, TPE and RSMLE procedures mentioned in this article to everybody who has access to a GAUSS licence with the optional package »Non-Linear Optimization«. We would like to thank Claude Hillinger, John Komlos, Albrecht Ritsehl, Christoph M. Schmidt, Ulrich Woitek, Klaus F. Zimmermann and the ECONHIST Research Group Munich for important comments and hints.

Prolog: Zum Indikator »Durchschnittliche Körpergröße«

Da der Indikator »Durchschnittliche Körpergröße« ein eher ungewöhnliches Maß für Ökonomen und Wirtschaftshistoriker ist, mögen zunächst einige einleitende Worte zum Aussagewert dieser Meßgröße angebracht sein.

Ein großer Teil der heutigen Wirtschaftshistoriker interessiert sich für die Frage, wie verschiedene Formen des Wirtschaftens bei unterschiedlichen Faktorausstattungen den Lebensstandard von Bevölkerungen oder sozialen Gruppen erhöht oder vermindert haben. Dabei ist kaum ein Begriff schwieriger zu definieren als »Lebensstandard«. Die gängige Praxis früherer Historiker-Generationen, ausschließlich Realeinkommen und Bruttosozialprodukte zu betrachten, ist vor allem in den letzten zwei Jahrzehnten heftig kritisiert worden: Man kann einfach keinen eindeutigen Zusammenhang zwischen Realeinkommen und anderen Komponenten des Lebensstandards, wie z.B. Gesundheit, Ernährungslage und Lebenserwartung, für alle Wirtschaftssysteme zumindest im historischen Verlauf annehmen¹. Es wurde vorgeschlagen, für diese nicht-materiellen Komponenten des Lebensstandards den Terminus »biologischer Lebensstandard« zu benutzen.²

Besonderes Interesse fand die Frage nach dem Zusammenhang von Ernährung und wirtschaftlicher Entwicklung. Ist eine quantitativ ausreichende und qualitativ hochwertige Ernährung die *Conditio sine qua non* für wirtschaftliche Entwicklung? Oder fanden gerade die Regionen und Länder zu neuen Formen des Wirtschaftens, die sich bei einer hohen und wachsenden Bevölkerungsdichte von Subsistenzkrisen bedroht fühlten (im Sinne einer »Not-macht-erfindersch«-Konzeption)?³ Zu derartigen Fragen kann die wirtschaftshistorische Forschung Einblicke vermitteln, die auch für gegenwärtige und zukünftige Entwicklungsprozesse relevant sein dürften.

Das größte Problem historischer Forschung besteht in der oft unzureichenden Quellenlage. Gerade der Faktor Ernährung ist hierfür ein gutes Beispiel: Eine alltägliche Angelegenheit wie die Nahrungsaufnahme wurde nicht gemessen und statistisch festgehalten. Um die Qualität der Ernährung im 18. und 19. Jahrhundert dennoch abschätzen zu können, benutzt die anthropometrische Forschung die durchschnittliche Körpergröße als Indikator. Dabei macht man sich die Tatsache zunutze, daß sich die Qualität der Ernährung in den ersten zwei Lebensjahrzehnten auf die Körpergröße auswirkt. Die Menge der eingenommenen Kalorien und Proteine wird dabei vom Verbrauch für Krankheiten und Arbeitsbelastung reduziert auf die Nettoernährung. Möchte man diese Nettoernährung messen, muß man freilich dem anderen wichtigen Einflußfaktor auf

¹ Eine exemplarische Diskussion findet man in Komlos (1993a), S. 7ff. Dessen Bedeutung für das Verhältnis von demographischer und ökonomischer Entwicklung wird in Komlos (1993c) erläutert.

² Ebd.

³ Vgl. Malthus (1798/1976), Boserup (1981), Komlos (1989).

die Körpergröße Rechnung tragen, dem genetischen Wachstumspotential. Zum einen verursacht die Genetik innerhalb einer Population eine Streuung der Größen, weswegen man ausschließlich die 'durchschnittliche' Körpergröße einer Population untersucht. Zum anderen existieren ethnologische Unterschiede in dieser durchschnittlichen Körpergröße. Um eine dadurch verursachte Verzerrung auszuschalten, muß man geringe Migration annehmen oder die Zusammensetzung der Samples kontrollieren, indem man z.B. Migranten gesondert analysiert (wenn dafür Informationen zur Verfügung stehen).

In der Praxis werden wir natürlich kaum die Daten einer gesamten Population zur Verfügung haben. Die mittlere Körpergröße muß also aus den überlieferten Stichproben geschätzt werden. Wie wir im folgenden sehen werden, vermitteln die Stichproben gerade für ältere Zeiträume wie das 18. Jahrhundert leider ein sehr verzerrtes Bild der Grundgesamtheit, so daß auch die üblichen Methoden der Varianz- u. Mittelwertschätzung¹ verzerrte Ergebnisse liefern. Wir werden deshalb in den folgenden Abschnitten

- definieren, was mit dem Problem der gestutzten bzw. genauer: der Verteilungen mit Shortfall gemeint ist,
- eine neuere Studie als Beispiel vorstellen, die aufgrund ungenügender Kenntnis der Schätzmethoden zu widersprüchlichen Ergebnissen kommt
- einen Überblick über die bisher bekannten Lösungsansätze für das Problem des Shortfalls geben,
- die Stärken und Schwächen dieser Lösungsansätze diskutieren und als Beispiel die vorläufigen Schätzungen vorstellen, die sich aus unseren Forschungen zur Ernährung im Bayern des 18. Jahrhunderts ergaben.

Im Zentrum der Betrachtung werden die Stärken und Schwächen der verschiedenen Schätzverfahren stehen, wobei herausgearbeitet wird, daß auch bestimmte Dateneigenschaften verstärkend auf die Schwachpunkte der Schätzer wirken können.

Als Anwendungsbeispiel für die dargestellten statistischen Methoden, werden in dieser Studie Schätzungen zur Entwicklung des Ernährungsstandards in Bayern zwischen 1750 und 1800 vorgestellt. Warum ist die Betrachtung dieser Entwicklung so interessant, daß sie einen derart großen statistischen Aufwand rechtfertigt?

Der geographische Raum des späteren Bayern erlebte genauso wie der größte Teil des übrigen Deutschlands im letzten Jahrtausend zwei fürchterliche Bevölkerungskrisen, im 14. und im 17. Jahrhundert. Wenngleich beide Krisen von exogenen Faktoren ausgelöst wurden (Beulenpest, Dreißigjähriger Krieg), kann man sich fragen, ob vielleicht die geringe Resistenz gegen diese Auslöser durch die jeweils zuvor stark angewachsene Bevölkerung, die angesichts relativ

¹ Gewöhnlich würde man für solch einfache Schätzungen die Momentenmethode verwenden, d.h. man setzt die entsprechenden Momente der Stichprobe mit der Grundgesamtheit gleich.

langsam wachsender Nahrungsmittelproduktion in sozialen Streß geriet, hervorgerufen wurde. Mit anderen Worten: Hatte Malthus Recht mit seiner pessimistischen Sicht der vor- und frühindustriellen Gesellschaft? Dabei ist zu bedenken, daß auch andere europäische Länder, wie etwa England im 17. Jahrhundert, in Bevölkerungskrisen gerieten, ohne daß dort ein derart furchtbarer Krieg wütete wie in Deutschland. Malthus verfaßte seine Gedanken Ende des 18. Jahrhunderts; wir wissen heute, daß die meisten Länder in den folgenden zwei Jahrhunderten von malthusianischen Krisen verschont blieben. Dennoch ist zu fragen, ob angesichts des Bevölkerungswachstums im 18. Jahrhundert gerade die wirtschaftlich weniger entwickelten Gebiete wie der bayerische Raum nur ganz knapp einer solche Krise entgingen: Wären die agrarischen Produktionssteigerungen im späten 18. und frühen 19. Jahrhundert geringer ausgefallen, hätte die Bevölkerung vielleicht epidemischen Krankheiten keine Widerstandskräfte entgegensetzen können.

Ein Indiz für eine drohende, aber noch nicht eintretende malthusianische Krise ist eine säkulare Abnahme des Ernährungsstandards. Der menschliche Körper stellt sich auf enger werdende Nahrungsspielräume ein, indem er zu geringerem Längenwachstum neigt. Dies setzt wiederum den erforderlichen Grundumsatz an Nahrungsmitteln (Basal Metabolic Rate) herab, der zur Lebenserhaltung benötigt wird. Eine deutliche Abnahme der durchschnittlichen Körpergröße wäre daher *ein* Hinweis auf eine drohende malthusianische Krise, wenn alle anderen wirksamen Faktoren kontrolliert worden sind. Die genauere Betrachtung dieser Faktoren (Krankheitsumfeld, Arbeitsbelastung, Selektivität der Stichprobenzusammensetzung) können in dieser Studie allerdings nicht untergebracht werden

1. Das Problem des »Shortfall«

Die wohl wichtigste Quelle für historische Körpergrößenmessungen stellen die militärischen Musterungslisten dar. Nur selten wurden jedoch komplette Geburtsjahrgänge für militärische Zwecke vermessen, wie es ausnahmsweise bei den bayerischen Conskriptionslisten des frühen 19. Jahrhunderts der Fall war. Aufgrund derart vollständiger Erfassungen ist allgemein akzeptiert, daß menschliche Körpergrößen auch in der Vergangenheit näherungsweise normalverteilt waren. In den meisten Ländern und Zeiträumen wurden nur die Männer in die Listen aufgenommen, die tatsächlich in einer Armee dienten, oder nur diese Listen überliefert. Die Rekrutierungsoffiziere nahmen jedoch fast ausschließlich solche Rekruten oder Freiwilligen in eine Armee auf, die ein bestimmtes Mindestmaß hatten und außerdem gesund und einigermaßen kräftig wirkten. Dies führt dazu, daß die kleineren Körpergrößen entweder überhaupt nicht oder mit geringer Fallzahl in der Verteilung repräsentiert sind. Dieser Fall hegt auch bei unserem bayerischen Sample für das 18. Jahrhundert vor.

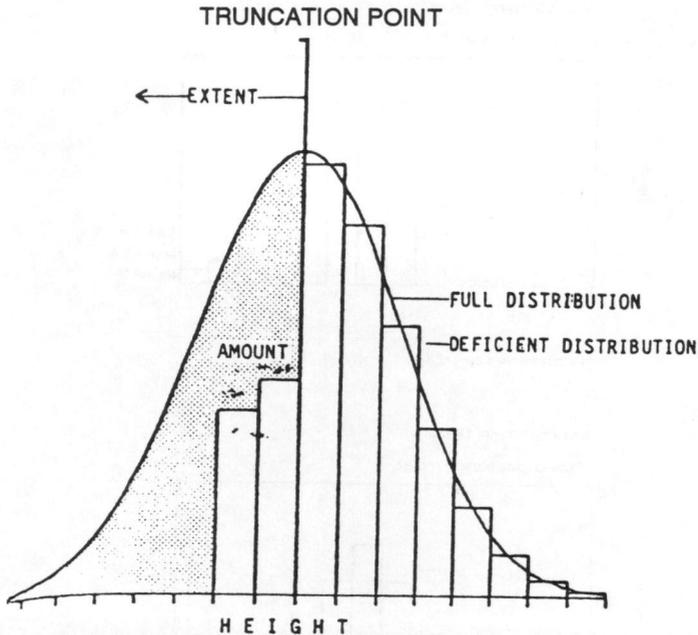


Abb. 1: Eine Verteilung von Körpergrößen mit Shortfall
Quelle: Wachter/Trussell (1982), S. 281.

Interessiert man sich nun z.B. für die Durchschnittsgröße der ausgewachsenen männlichen Bevölkerung und schätzt man diese mit obigen Daten, wird der Schätzer immer nach oben verzerrt sein. Wäre die Stichprobe an einem bestimmten Punkt (*Truncation Point*, nachfolgend: t.p.) vollständig abgeschnitten, würde dies dennoch keine Probleme bereiten, ließe sich doch der Schätzer mit Hilfe der vorhandenen Methoden zur Analyse von *truncated samples* entzerren.¹

Leider muß man dagegen in der Anthropometrie von Stichprobenverteilungen ausgehen, die nicht vollständig abgeschnitten sind, sondern noch links eines bestimmten Punktes - nennen wir ihn wieder *Truncation Point* (tp.) - eine, allerdings überproportional geringe Dichte besitzen. Dies kommt nicht zuletzt auch daher, daß das Mindestmaß oft nur ein Richtmaß war, nach dem sehr willkürlich verfahren wurde. Außerdem wurden Rekruten, die besonders kräftig wirkten oder für bestimmte Zwecke gebraucht wurden (z.B. Chirurgen, Fahnenträger), auch mit einer sehr geringen Körpergröße akzeptiert.

Zusammenfassend muß man feststellen, daß in unseren Stichproben Daten »fehlen«, wenn man Inferenzen über gewisse Populationen machen möchte .

¹ Vgl. Maddala (1983), S. 165-169; siehe auch den Abschnitt 3.2 in diesem Beitrag.

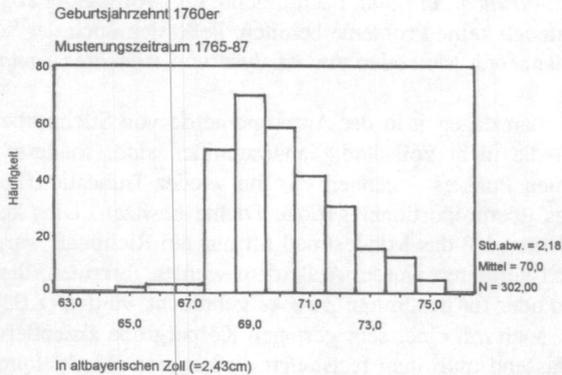
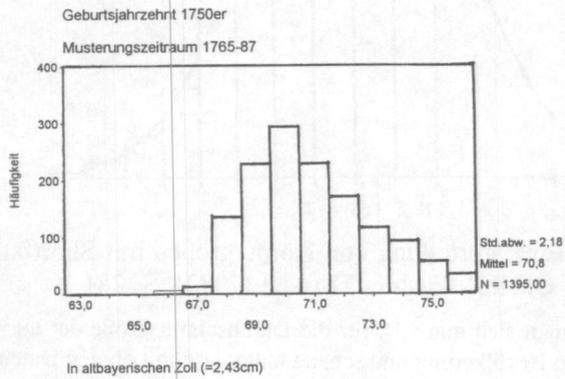
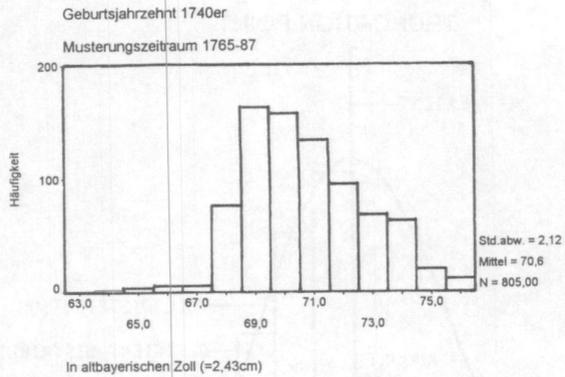


Abb. 2a: Verteilungshistogramme des Samples "Bayerische Soldaten des späten 18. Jahrhunderts", Musterungszeitraum 1765-87
Quelle: Eigene Erhebung aus dem Bayer. Kriegsarchiv, Bestand AV14d

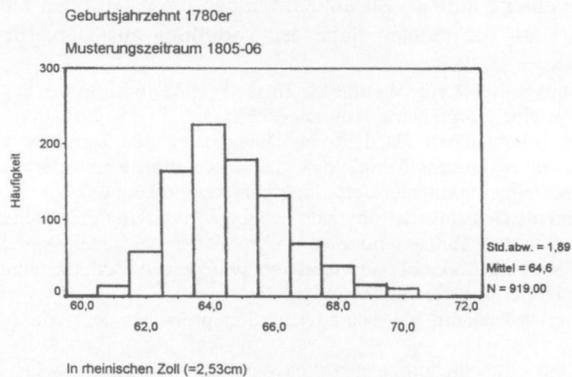
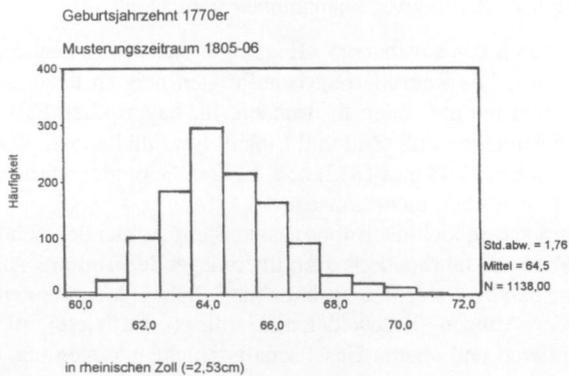
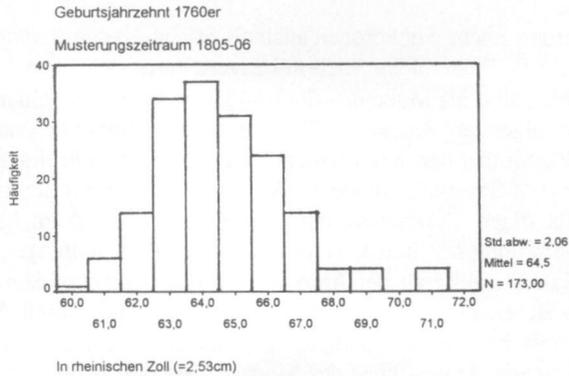


Abb. 2b: Verteilungshistogramme des Samples "Bayerische Soldaten des späten 18. Jahrhunderts", Musterungszeitraum 1805-06
Quelle: Eigene Erhebung aus dem Bayer. Kriegsarchiv, Bestand AV14d

Deswegen werden solche Stichproben auch als Stichproben mit *Shortfall* (Fehlbetrag) bezeichnet. Eine solche Stichprobenverteilung sieht man in Abb. 1.^{6,7} Zu unterscheiden sind als Maße des Shortfalls der Anteil der fehlenden Beobachtungen (im folgenden: *Amount* of Shortfall, in der Abbildung grau) von dem Bereich der Stichprobe, der durch fehlende Beobachtungen beeinträchtigt wird (der sog. *Extent* of Shortfall). Beide Maße werden in Prozent der zugrundeliegenden, vollständigen Gesamtverteilung ausgedrückt.⁸ Weil im beschädigten Bereich der Verteilung fast immer zumindest einige Beobachtungen vorhanden sind, ist der Extent größer als der Amount.⁹ Daß der Extent of Shortfall in der Abbildung im Bereich des Modalwertes beginnt ist reiner Zufall. In den Verteilungen unseres bayerischen Samples beginnt er - nach dem graphischen Eindruck zu urteilen - meist links des Modalwertes (vgl. Abb. 2a u. 2b). Weitere Abweichungen von der üblichen Normalverteilung werden in vielen historischen Samples oft von zwei Phänomenen verursacht:

- Zum einen durch das sogenannte »Heaping«, das von ungenauen Meßmethoden herrührt: Die Rekrutierungsbeauftragten neigten dazu, zu den größeren Einheiten hin auf- oder abzurunden. Im bayerischen Fall gab es die historischen Einheiten Fuß, Zoll und Linien. Ein Fuß hatte 12 Zoll, ein Zoll 12 oder (zwischen 1788 und 1812) acht Linien. Gerundet wurde oft auf den ganzen Zoll, aber eben nicht konsequent.
- Zum anderen setzen sich die militärischen Samples aus der Zeit des späten 18. und frühen 19. Jahrhunderts zum überwiegenden Teil aus Angehörigen der unteren sozialen Schichten zusammen. Es gibt jedoch kleinere Gruppen innerhalb der Armeen (Elitesoldaten, Gardisten, Offiziere), die teilweise aus den mittleren und oberen Gesellschaftsschichten kamen und daher besser ernährt und größer waren. Hier muß darauf geachtet werden, daß kein Sampling-Fehler¹⁰ auftritt. Zusätzlich können diese Gruppen einen schwachen Peak¹¹ auf der rechten Seite der Verteilung auslösen. Im folgenden

⁶ Auch in diesem Beispiel von Wächter & Trussell (1982) wird in der zugrundeliegenden Population eine Normalverteilung angenommen.

⁷ Im folgenden benutzen wir als deutsche Übersetzung den Terminus »beschädigte Verteilung«, weil »gestutzte Verteilung« die Assoziation einer klar an einem bestimmten Punkt abgeschnittenen Verteilung hervorrufen könnte.

⁸ Natürlich kann als Gesamtverteilung jede beliebige Verteilung angenommen werden. Doch in den meisten Fällen wird man aufgrund der bekannten Verteilungseigenschaften von Körpergrößen oder aufgrund von zwingenden Modellannahmen (s. Kap. 3) die Normalverteilung verwenden.

⁹ Im Falle einer vollständig abgeschnittenen Stichprobe ist der Extent gleich dem Amount.

¹⁰ Es ist klar, daß eine Stichprobe natürlich möglichst repräsentativ für die entsprechende Grundgesamtheit gezogen werden sollte. Deswegen muß man auch vermeiden, aus den rechten Gruppen überdurchschnittlich viele Fälle zu erfassen. Dies könnte vorkommen, da sie oft am Anfang der Listen aufgeführt werden oder aufgrund ihrer Schichtzugehörigkeit interessanter für soziale Vergleiche sind.

¹¹ I.e. eine unerwartet hohe Dichte in einem eng umgrenzten Bereich der Verteilung.

werden diese Gruppen aufgrund ihrer Position innerhalb der Gesamtverteilung als »rechte Gruppen« bezeichnet.

Dennoch ist es vor allem der Shortfall, der es schwer macht, die Bevölkerungsgruppen, aus denen diese Soldaten stammen, zu untersuchen, und das nicht nur bezüglich der Schätzung von Mittelwert und Standardabweichung, sondern vor allem auch bei anspruchsvolleren Analysen wie z.B. bezüglich des Einflusses exogener Variablen (i.e. Regressionsrechnung). Im Laufe der Zeit wurden dazu eine Reihe von Verfahren entwickelt, die fast alle ihre Stärken und Schwächen haben. Hierzu gehören das Quantile Bend Estimate (QBE), das Reduced Sample Maximum Likelihood Estimate (RSMLE) und die Komlos&Kim-Methode (K&K).

2. Von der Notwendigkeit, die Eigenschaften dieser Schätzmethoden zu kennen: Das Beispiel der württembergischen Trends im späten 19. Jahrhundert

Nicht zuletzt aufgrund der zahlreichen Gefahren, die durch Verzerrungsfaktoren gegeben sind, hat sich in der anthropometrischen Forschung eine Replikationskultur entwickelt.¹² Wie in der naturwissenschaftlichen Forschung werden (im Idealfall) die benutzten Samples nach der Erstveröffentlichung der Ergebnisse kostenlos zur Verfügung gestellt, damit andere Wissenschaftler mit anderen Schätzverfahren möglicherweise genauere oder auch andere Ergebnisse präsentieren können. Ein Beispiel, bei dem die Ergebnisse von drei Schätzverfahren gleich gemeinsam veröffentlicht wurden, ist Sophia Twarogs Studie zu Lebensstandards in Württemberg in der zweiten Hälfte des 19. Jahrhunderts.¹³ Ihre Dissertation ist die erste anthropometrische Studie zur deutschen Wirtschaftsgeschichte, die eine Stichprobe aus dem militärischen Bereich benutzt. Sophia Twarog stellt sich die Frage, ob diese Phase der Industriellen Revolution negative Auswirkungen auf die Ernährung der Bevölkerung gehabt hat. Sie kommt zu dem Ergebnis, daß diese These aufgrund der Einbrüche in den frühen 1870er und mittleren 1880er Jahren bejaht werden kann. (Abb. 3a und 3b). Sie stützt sich dabei auf die Ergebnisse des QBE, während sie bei der Schätzung mit den beiden anderen Verfahren zu ganz anderen Ergebnissen kommt. Die methodischen Ursachen dieser Abweichungen werden leider nicht ausreichend diskutiert. Neben Problemen des Samplings¹⁴ und der exogenen Variablen¹⁵ scheinen uns die Unterschiede vor allem aus den Eigenschaften der Schätzer herzuführen.

¹² Vgl. z.B. Komlos (1993b).

¹³ Twarog (1993).

¹⁴ In ihrem Sample ist die Gruppe der Einjährig-Freiwilligen, welche ungefähr der Oberschicht entspricht, unserer Einschätzung nach um ein Vielfaches überrepräsentiert.

¹⁵ Die Soldaten entstammen verschiedenen Altersgruppen. Twarog berücksichtigt dies

3. Die Funktionsweise der Schätzverfahren

3.1 Das Quantile Bend Estimate (QBE)

Das QBE wurde 1981 von Wächter & Trussell (1982) speziell für die Schätzung von Mittelwert und Standardabweichung von Körpergrößen-Stichproben mit Shortfall entwickelt. Ihre QBE-Schätzungen erwiesen sich bei ihren eigenen Modellrechnungen für zehn verschiedene Verteilungstypen als relativ robust und effizient. So wurde diese Prozedur im vergangenen Jahrzehnt besonders häufig verwendet, wenngleich sich in der Praxis gezeigt hat, daß z.B. mit QBE geschätzte Trends von Mittelwerten meist unvernünftig stark schwanken.¹⁶

Die Funktionsweise des QBE¹⁷ besteht darin, daß er sich denjenigen Amount of Shortfall sucht, der die Stichprobenverteilung am besten zu einer Normalverteilung ergänzt. Dabei muß man allerdings annehmen, daß der obere Teil der Verteilung, also der Teil rechts eines (möglichen) Truncation Points, völlig unbeschädigt ist. Die Ermittlung des besten Amounts A^* geschieht mit Hilfe von Quantile-Quantile-Plots¹⁸, wobei für jede Körpergröße Y auf der Y-Achse ein X berechnet wird gemäß

$$X = \Phi^{-1}((1-A)(1-F(Y))), \quad (1)$$

mit dem entsprechenden Amount A ,
der $N(0,1)$ -Verteilungsfunktion Φ ,
und der empirischen Verteilungsfunktion F .

Je »besser« der Amount ist, desto gerader muß der QQ-Plot rechts des dazugehörigen Truncation Points verlaufen. Die Güte dieser »Geradheit« wird durch eine robuste Regression ermittelt, also eine Art Weighted-Least-Square Regression, die den aussagekräftigeren oberen Bereich der Verteilung stärker und den unteren Bereich schwächer und schließlich gar nicht mehr gewichtet.

allerdings nicht: Sie müßte aber die Stichproben entsprechend unterteilen oder exogenen Variablen einführen, die den möglichen Größenunterschied etwa von 19- und 23-jährigen kontrollieren.

¹⁶ Komlos & Kim (1990).

¹⁷ Die Funktionsweise des QBE ist nichttrivial und bedarf zum vollen Verständnis fundierter statistischer Kenntnisse. Der interessierte Leser sei deswegen für eine genaue Darstellung auf den Originalartikel von Wächter & Trussell (1982) verwiesen.

¹⁸ Quantile-Quantile-Plots dienen zum graphischen zweier Verteilungen. Dabei werden die Quantile der einen Verteilung bezüglich der Quantile der anderen Verteilung geplottet. Stimmen die Verteilungen überein, ergibt sich im Plot eine Gerade. Stimmen sie z.B. nur oberhalb eines bestimmten Quantils überein, wird man zwar auch in diesem Bereich eine gerade Linie registrieren, anderswo aber eine Abweichung von der Geraden erkennen können. Statistikpakete wie SPSS und SAS bieten dieses einfache, aber nicht uneffektive deskriptive Verfahren zum nichtparametrischen Vergleich zweier Verteilungen an.

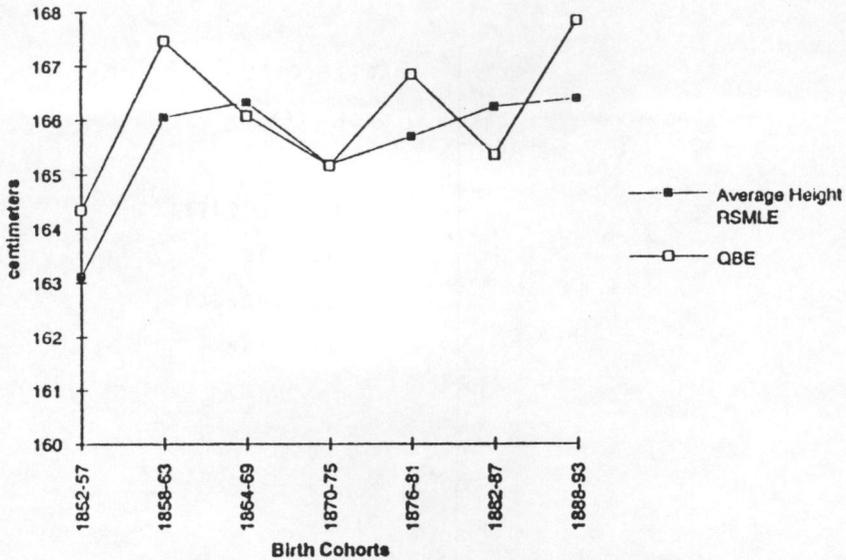


Abb. 3a: Schätzungen der Körpergrößen von württembergischen Soldaten im späten 19. Jahrhundert mit QBE und RSMLE, 7 Zeitabschnitte à 6 Jahren
 Quelle: Twarog (1993), S. 302.

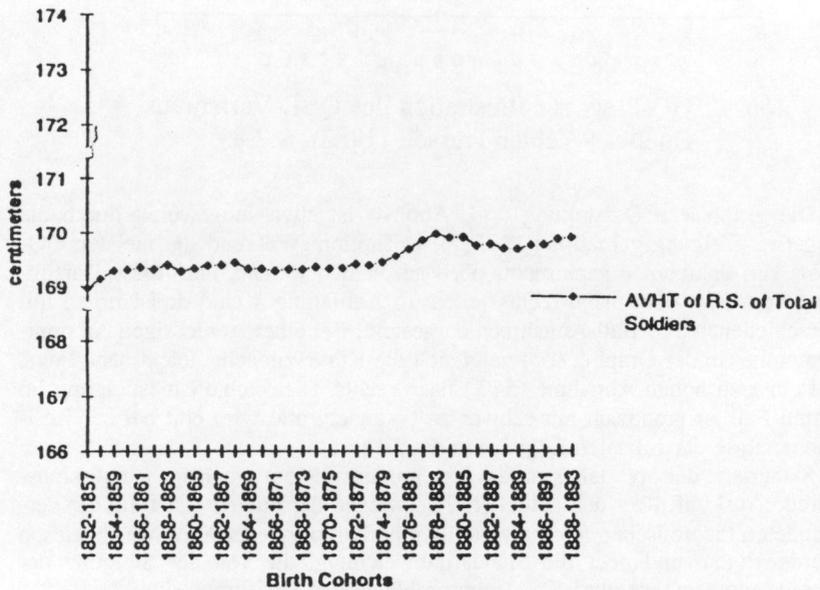


Abb. 3b: Schätzungen der Körpergrößen von württembergischen Soldaten im späten 19. Jahrhundert mit dem K&K-Verfahren, Gleitende 5-Jahres-Mittelwerte
 Quelle: Twarog (1993), S. 305.

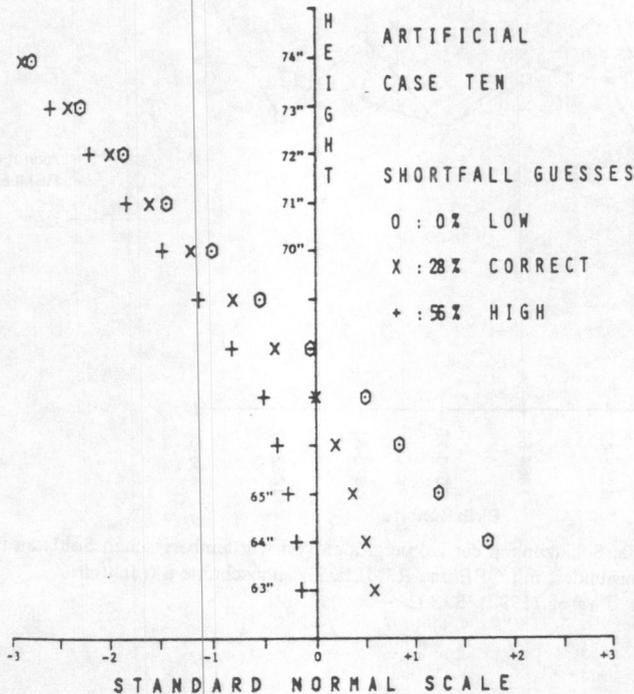


Abb. 4: QQ-Plots zur Illustration des QBE-Verfahrens
Quelle: Wachter/Trussell (1982), S. 283.

Die graphische Darstellung (vgl. Abb. 4) ist etwas ungewohnt durch die negative Steigung gemäß der $(1-F(y))$ -Definition: Während die meisten QQ-Plots von links unten nach rechts oben gezeichnet werden, läuft diese Darstellung von links oben nach rechts unten. In Abbildung 4 sind drei Kurven mit verschiedenen Shortfall-Annahmen dargestellt. Bei einer zu niedrigen Amount-Annahme (in der Graphik 0%) neigt sich die Kurve zunächst leicht nach links, bei einer zu hohen Annahme (54%) nach rechts. Die leichte Linksneigung im ersten Fall ist graphisch nur schwer zu erkennen, und wird erst bei der Residuenanalyse der **robusten Regression** deutlich.

Steht dann der optimale Amount fest, gibt uns die dazugehörige Regressionsgerade Auskunft über den Mittelwert, der wegen der Standardisierung der verwendeten theoretischen Normalverteilung im Nullpunkt der X-Achse abgelesen werden kann, und über die Standardabweichung, die von der Steigung der Gerade repräsentiert wird. Die Hauptprobleme des Verfahrens sind:

- die extrem starke Abhängigkeit von der Normalverteilung¹⁹ und

¹⁹ Dieses Manko war Wachter & Trussell (1982) durchaus bewußt. Deshalb rieten sie von der Benutzung des QBE in anderen Zusammenhängen eher ab.

- die Verwendung von robusten Regressionen, deren Gewichtungsfaktoren von sehr willkürlichen Annahmen abhängen.²⁰

3.2 Das Reduced-Sample-Maximum-Likelihood Verfahren (RSMLE)

Das zweite Verfahren für die Schätzung von Mittelwert und Standardabweichung ist nicht speziell für die Körpergrößenforschung entwickelt worden, sondern basiert auf den bekannten Methoden zur Analyse von *truncated distributions*.²¹ Das Sample wird am Truncation Point klar abgeschnitten und eine Maximum-Likelihood-Schätzung für den verbleibenden, unbeschädigten Bereich der Verteilung vorgenommen, um Mittelwert und Varianz bzw. Standardabweichung zu erhalten. Die Dichtefunktion einer links des Punktes a abgeschnittenen normalverteilten Zufallsgröße X mit Mittelwert μ und Standardabweichung σ lautet²²

$$f(X) = \frac{\sigma^{-1} f_{\Phi}((X - \mu)/\sigma)}{1 - \Phi((a - \mu)/\sigma)}, \quad (2)$$

wobei f_{Φ} die Dichtefunktion der Standardnormalverteilung ist.

Das zentrale Problem ist es jetzt, den Truncation Point a optimal zu schätzen. Man hat hier einen Trade-off: Wird ein a zu weit links gewählt, ist der Schätzer verzerrt, weil ein Teil der Beschädigung in der Verteilung in die Schätzung eingeht. Wird dagegen der t.p. zu weit rechts angesetzt, geht unnötig viel Information verloren. Gerade wenn der t.p. ganz nah links vom Mittelwert liegt, wird die Schätzung relativ ineffizient. Falls er gar rechts vom Mittelwert liegen sollte, dürfte die Schätzung keinen Sinn machen. Die Varianz der Schätzungen steigt zu stark an. Zur Bestimmung des t.p. wurden bisher drei Verfahren vorgeschlagen:

- Vorinformation: Zwar gibt es meist Vorschriften für die Rekrutierungsverfahren, in denen das Mindestmaß angegeben ist. Die empirischen Befunde deuten jedoch darauf hin, da diese Vorschriften gerade im 18. Jahrhundert nicht exakt eingehalten wurden. Der wahre Truncation Point kann sowohl über als auch unter dem gesetzlichen Mindestmaß gelegen haben.
- Die Methodik von Wachter & Trussell (1982). Wachter & Trussell (1982) schlagen vor, eine Sequenz von Schätzungen z.B. des Mittelwertes vorzunehmen. Ihr Verfahren schneidet auf linken, beschädigten Seite Zoll für

²⁰ S. auch Greene (1993)

²¹ »Truncated distributions« sind Verteilungen, die unterhalb eines klar definierten Truncation Points keine Beobachtungen mehr aufweisen, also tatsächlich abgeschnitten sind.

²² Einen verständlichen Einblick in die Schätzung von truncated distributions vermittelt Greene (1993), S. 682 ff. Tieferegreifende Analysen findet man in Maddala (1983).

Zoll von der Verteilung ab und nimmt jedes Mal für den Rest der Verteilung eine Schätzung vor. Sobald die Schätzungen sich nicht mehr stark unterscheiden, weiß man, daß man den Truncation Point erreicht bzw. gerade passiert hat. Weil man bei historischen Samples nicht erwarten kann, daß die Schätzungen exakt gleich sind, muß der Forscher einen Bereich festlegen, wie weit die Schätzungen auseinanderliegen dürfen. Von drei Schätzungen (beim t.p., davor und danach) wird daraufhin der Median als plausibelste Schätzung ausgewählt. Ein zweiter Vorschlag von Wachter & Trussell (1982) arbeitet mit der Varianz der Schätzungen des Mittelwerts²³. Man gibt sich eine obere Grenze für diese Varianz vor (d.h. man versucht den Informationsverlust zu begrenzen), und schneidet die Stichprobe dann an der Stelle ab, an der diese Varianz gerade noch eingehalten werden kann. Das Problem beider Vorschläge ist allerdings wieder die Willkür, hier die Willkür bei der Festsetzung des Streubereichs des Schätzers.

- Parameterfreie Methoden. Heintel (1995) schlägt die Verwendung von Kerndichteschätzern vor, um den t.p. zu ermitteln. Die zu lösende Aufgabe besteht darin, Heaping und zufällige Abweichungen der empirischen von einer erwarteten, »glatteren« Verteilung²⁴ (im unbeschädigten Bereich) zu ignorieren und nur den t.p. zu erkennen: Es ist zu erwarten, daß direkt links des t.p. die Dichte des Samples im Vergleich zum Bereich rechts des t.p. überproportional gering wird.

Für die Beobachtungen X_i werden innerhalb der Beobachtungsspannweite äquidistante Punkte x erzeugt, an denen eine »Dichte« $\hat{f}_h(x)$ folgendermaßen geschätzt wird:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (3)$$

mit der Stichprobenlänge n ,
 der Bandbreite h ,
 und einer nichtnegativen Gewichtsfunktion K .

Als Gewichtsfunktionen stehen verschiedene Möglichkeiten zur Verfügung, deren Eigenschaften einander relativ ähnlich sind: Quartic, Triangle, usw.²⁵ Die entsprechende Bandbreite kann aufgrund einer von Härdle (1990) über die Minimierung des *Mean Integrated Square Error* (MISE) ermittelten Tabelle, gewählt werden.²⁶ Mit Hilfe der absoluten Differenzen von (3)

²³ Natürlich lassen sich beide Verfahren auch bezüglich der Schätzung der Standardabweichung durchführen. Doch in den meisten Fällen ist man eher an einer robusten Mittelwertschätzung interessiert.

²⁴ Dieses Verfahren hängt nicht von einer expliziten Verteilungsannahme ab.

²⁵ Eine gute Einführung in die Kerndichteschätzung findet man in Härdle (1990).

²⁶ Ebd., S. 76.

erhält man dann einen Schätzer für den Truncation Point.²⁷ Mit diesem TPE (*truncation point estimate*) kann anschließend die RSMLE-Schätzung von Mittelwert und Varianz durchgeführt werden.

3.3 Die Komlos & Kim-Methode

Untersucht man die Entwicklung von Körpergrößen in der Zeit, interessiert oft nicht die absolute durchschnittliche Körpergröße, sondern nur die relative Änderung der Körpergröße in der Zeit, oder noch einfacher nur das Vorzeichen der Änderung. Die schon erwähnten starken Schwankungen der mit QBE geschätzten Mittelwerte, die leider auch beim RSMLE mit t.p.-Ermittlung über Vorinformation oder das Wachter & Trussell-Verfahren zu beobachten sind, veranlaßten Komlos & Kim (1992), eine robusteres Verfahren einzuführen, das nur die Richtung des Trends schätzt. Dabei werden einfach alle Stichproben-Zeitreihen an einem einheitlichen t.p. abgeschnitten, und die Mittelwerte der gestutzten Stichproben berechnet. Es ist klar, daß dieser t.p. mindestens so groß sein muß wie der t.p. dieser Stichprobenverteilung, die am weitesten nach oben beschädigt ist, aber in keinem Fall größer sein sollte.²⁸

Gerade für die »ersten Eindrücke« ist die K&K-Methode gut geeignet, man muß sich allerdings vergegenwärtigen, daß neben der schon erwähnten anderen Niveaus der Schätzer, noch zwei andere Probleme die Effizienz dieser Methode beeinflussen können:

- Die Varianz muß bei allen Stichproben gleich und gegeben sein.
- Bei in der Zeit stark »wandernden« Truncation Points, kann durch die Wahl des größten t.p.'s bei einigen Stichproben extrem viel Information verlorengelassen werden.

Eine andere Methode zur Trendschätzung wäre die Verwendung der Modalwerte der Stichproben. Bei der (theoretischen) Normalverteilung sind bekanntlich Mittelwert und Modalwert gleich. Bei einer ursprünglich approximativ normalverteilten Stichprobe, die von Shortfall heimgesucht wurde, kann man daher davon ausgehen, daß ihr Modalwert mit dem Mittelwert der unbekanntenen »vollständigen« Stichprobe ungefähr übereinstimmt, wenn der Shortfall nicht über diesen unbekanntenen Mittelwert hinaus übergreifen hat. Mit Hilfe der schon erwähnten Kerndichteschätzer, macht es keine Probleme, den Modalwert einer Stichprobe zu bestimmen. Ist man zusätzlich noch an statistischen Tests interessiert, wäre dies z.B. über Bootstrap-Methoden zu erreichen²⁹. Gerade hier kann sicherlich noch einige Forschungsarbeit investiert werden.

²⁷Heintel (1995) liefert eine genaue Beschreibung dieser Vorgehensweise, incl. eines für jedermann leicht nachvollziehbaren (und nachprogrammierbaren) Algorithmus.

²⁸ Eine sinnvolle Methode wäre es z.B., mit der Heintel-Methode alle t.p.'s zu schätzen, und dann den größten t.p. auszuwählen.

²⁹ S. Efron (1982).

Tab. 1: Wichtigste Stärken und Schwächen der Schätzverfahren (Mittelwert oder Trendrichtung bei gruppierten Körpergröße-Samples mit Shortfall)

	Schätzgenauigkeit	Annahmen, Voraussetzungen (generell: Norm.vert.)	Exogene Variablen
QBE	große Varianz der Schätzungen. Mittelw., Std.abw.	besonders empfindlich für "rechte Gruppen" (s. Kap. 1)	nur durch Sampleteilung
RSMLE	genauester absoluter Schätzer Mittelw., Std.abw. Vorsicht bei Heaping	t.p. nicht zu nah an Mittelwert	schätzbar und damit kontrollierbar
K&K	robust für Trendschätzung	Ann. konstanter Varianz. Große Fallzahl.	nur durch Sampleteilung

4. Erfahrungen mit den einzelnen Schätzverfahren

Nachdem zuvor die verfügbaren Methoden charakterisiert wurden, sollen im folgenden Abschnitt ihre Eigenschaften empirisch verglichen werden. Einen Überblick über die nachfolgend erläuterten Ergebnisse gibt Tabelle 1. Während für den QBE-RSMLE-Vergleich bereits auf eine Simulationsstudie zurückgegriffen werden kann, werden RSMLE und die K&K-Methode anhand von zwei Beispielen verglichen. Neben Twarogs bereits erwähnter Studie wird eine Trendschätzung für Bayern im 18. **Jahrhundert** herangezogen.

Wie bereits erwähnt, basiert das QBE stark auf der Normalverteilung. Man wird erwarten, daß dadurch der Schätzer schon bei kleineren Abweichungen von der Normalverteilung, wie sie z.B. durch die rechten Gruppen auftreten können, ungenaue Ergebnisse liefert. Doch selbst bei normalverteilten Stichproben bleibt die Varianz der Schätzungen sehr hoch, wie Heintel (1995) anhand einer Simulationsstudie feststellte. Neben der Tendenz des QBE, den wahren Mittelwert zu unterschätzen, erwies sich die große Varianz der Schätzungen **als** besorgniserregend. Während die Varianz bei RSMLE kombiniert mit TPE (TPE/RSMLE) bei den jeweils 1000 Schätzungen immer geringer als 1cm^2 war, lag sie beim QBE **nie unter** 3cm^2 . Es ist zu erwarten, daß bei einer **empirischen**

Verteilung mit z.B. einer leicht verzerrten rechten Seite die Ineffizienz des Schätzers noch größer sein dürfte.

Die erwähnte Simulationsstudie zeigte, daß der RSMLE zumindest in Kombination mit der Heintel-Methode gegenüber dem QBE Effizienzvorteile hat. Auf der anderen Seite erschien zumindest eine RSMLE-Schätzung in der Studie von Twarog als unplausibel (vgl. Abb. 3a u. 3b): Warum sollten in den frühen 1850er Jahren innerhalb eines Jahrünfts die durchschnittlichen Körpergrößen um rund 3cm zunehmen? Dies wäre eine historisch einzigartige Verbesserung der Ernährungslage in einem so kurzen Zeitraum. Zudem bewegt sich der K&K-Schätzer kaum. Das Problem liegt wohl in der Festlegung des Truncation Points: Twarog verwendete die Vorinformation, daß das Mindestmaß bei 162cm lag und setzte selbst aus Vorsicht 162,9cm an. Dadurch liegt der t.p. in diesen Stichproben sehr hoch. Im konkreten Fall war er fast identisch mit dem geschätzten Mittelwert. Je kleiner aber der Bereich der Verteilung wird, der in die Schätzung eingeht, desto ungenauer wird der RSMLE. Leider stehen die Daten noch nicht zur Replikation zur Verfügung, so daß die Schätzergebnisse bis jetzt nicht mit denen anderer Schätzmethoden (wie z.B. der Heintel-Methode) verglichen werden konnten. Inhaltlich wirken die Ergebnisse der K&K-Schätzung dieses Zeitraumes plausibler³⁰, weil es nach den bekannten Begleitumständen dieser Zeit trotz der besseren Ernten keinen überzeugenden Grund für eine derart enorme Verbesserung der Ernährungssituation gab.

In unseren exemplarischen Schätzungen des bayerischen Ernährungsstandards im späten 18. Jahrhundert verwendeten wir aufgrund der Ungenauigkeit des QBE nur den K&K-Schätzer sowie den TPE/RSMLE (Abb. 5 und Tabelle 2)³¹. Für den K&K-Trendschätzer legten wir zum einen das höchste Mindestmaß zugrunde (167,8cm), zum anderen den höchsten »praktizierten« Truncation Point, wie ihn der TPE ermittelt (165,2cm). Es erscheint uns historisch durchaus plausibel, daß in der Musterungs-Zeitperiode von 1765-87 (Geburtsjahrzehnte 1740er-60er), in der das vorschriftsmäßige Mindestmaß äußerst hoch lag, auch sehr viele Soldaten unterhalb dieses Maßes akzeptiert wurden. Eine visuelle Analyse der Stichprobenverteilung dieser Zeit weist in die gleiche Richtung: Beim Geburtsjahrzehnt der 1760er sind z.B. links des vorschriftsmäßigen Mindestmaßes von 69 Zoll noch ausgesprochen viele Beobachtungen vorhanden (vgl. Abb. 2a, unteres Diagramm).

Ein Vergleich der drei Schätzungen für Bayern zeigt eine weitgehende Übereinstimmung der Bewegungsrichtungen (vgl. Abb. 5). Die Schätzungen mit TPE/RSMLE schwanken nicht unvernünftig hoch, verhalten sich also wie die einfachen K&K-Trendschätzungen, die sich in der Forschungspraxis, z.B. auch in Körpergröße-Studien in Schulen und Gefängnissen immer wieder als

³⁰ An dieser Stelle sei noch einmal erwähnt, daß sich die K&K-Schätzer aufgrund des vollständigen Abschneidens der Stichprobe auf einem anderen, höheren Niveau bewegen.

³¹ Abb. 5 zeigt die Werte der Tabelle 2 in graphischer Form, mit Ausnahme des TPE.

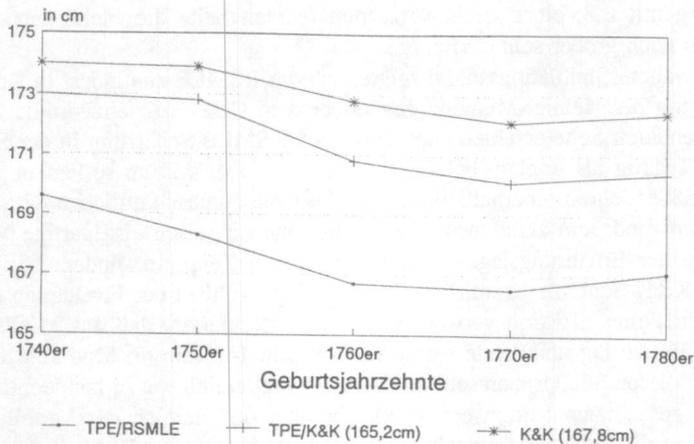


Abb. 5: Geschätzte Trends der Körpergrößen von bayerischen Soldaten (verschiedene Schätzverfahren)
(Quelle: wie Abb. 2a/b)

Tab. 2 Schätzungen der durchschnittlichen Körpergrößen von bayerischen Soldaten im späten 18. Jahrhundert mit verschiedenen Verfahren
(Quelle: Wie Abb. 2a/b. Geburtsjahrzehnt 1740er-1760er aus den Musterungen 1765-87; Geburtsjahrzehnt 1770er/80er aus den Musterungen 1805/06)

Geburts- jahrz.	N	TPE	TPE/ RSMLE	TPE/K&K (165,2cm)	MM/K&K (167,8cm)
1740er	805	165,1	169,6	173,0	174,0
1750er	1306	165,3	168,3	172,8	173,9
1760er	302	164,3	166,8	170,8	172,8
1770er	1138	165,2	166,5	170,1	172,1
1780er	919	162,8	167,1	170,4	172,4

Abkürzungen

N Zahl d. Fälle

TPE Truncation Point Estimate (Heintel-Schätzer)

TPE/RSMLE auf letzterem basierendes Reduced-Sample-Maximum-Likelihood-Estimate

TPE/K&K () auf dem TPE basierende K&K-Schätzung (alle Beobachtungen unter 165,2cm wurden entfernt.)

MM/K&K () Auf dem höchsten verwendeten Mindestmaß basierende K&K-Schätzung (alle Beobachtungen unter 167,8cm wurden entfernt.)

überraschend robust erwiesen haben. Ist man hauptsächlich an Trendentwicklungen interessiert, halten wir daher eine Doppelschätzung von K&K und TPE/RSMLE für die zuverlässigste Methode, weil man die Schätzer auf diese Weise wechselseitig kontrollieren kann.³²

Der große Vorteil der RSMLE-Methode ist vor allem darin zu sehen, daß nach der Ermittlung des t.p.'s, auch exogene Variablen in Form einer *truncated regression* mitmodelliert werden können, die z.B. Unterschiede in der sozialen, regionalen oder der schon erwähnten altersmäßigen Zusammensetzung der Stichprobe kontrollieren helfen. Darauf kann natürlich in diesem Beitrag nicht mehr eingegangen werden.

Zusammenfassung

Für das Problem des Shortfall bei der Nutzung des Indikators »Durchschnittliche Körpergröße« wurden in den vergangenen 12 Jahren einige Lösungsvorschläge gemacht. Leider produzieren die verschiedenen Schätzer nach wie vor oft unterschiedliche Ergebnisse. In diesem Beitrag wurde zunächst ein Überblick über die prinzipielle Arbeitsweise der unterschiedlichen Schätzer gegeben. Die zentrale Frage war: Wie kommt es zu den unterschiedlichen Schätzergebnissen? Die QBEMethode erscheint aufgrund der Simulationsergebnisse von Heintel (1995) und der Volatilität der bisher veröffentlichten Schätzungen als ineffizient. Bei der RSMLE-Methode ist es wichtig, ein Verfahren zur Bestimmung des Truncation Points anzuwenden, das einerseits nicht zu viele Fälle von der Analyse ausschließt, andererseits aber keinen t.p. innerhalb des Shortfall-Bereichs der Stichprobe auswählt. Bis jetzt scheint das am besten geeignete Verfahren die von Heintel (1995) vorgestellte Methode (TPE) zu sein, die natürlich in der Praxis noch genauer überprüft werden muß. Für reine Trendanalysen dürfte die robuste Komlos&Kim-Methode in Kombination mit TPE/RSMLE die sicherste Methode sein, mit dem Problem des Shortfalls umzugehen. Zur Überprüfung dieser Methoden wurden in der Studie auch erste Schätzergebnisse für Trends der bayerischen Körpergrößenentwicklung vorgelegt, die auf eine Verschlechterung der Ernährungsstandards im späten 18. Jahrhundert hindeuten.

³² Die Kontrolle des TPE/RSMLE ist vor allem deshalb vernünftig, weil der TPE erst neu entwickelt wurde und sein Verhalten in »Extremisituationen« wie extremen heaping oder truncation points, die nahe beim Mittelwert liegen, noch nicht genau bekannt ist. Die Kontrolle der K&K-Schätzungen ist auf jeden Fall empfehlenswert, da diese Methode - wie schon erwähnt - von sehr restriktiven Annahmen ausgeht.

Literatur

- Boserup, E. (1981), *Population and Technological Change: A Study of Long-Term Trends*. Chicago: University of Chicago Press.
- Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Siam Publishing.
- Greene, W.H. (1993), *Econometric Analysis*. New York: MacMillan.
- Härdle, W. (1990), *Smoothing Techniques*. New York: Springer.
- Heintel, M. (1995), *A Simple Method for Identifying the Truncation Point of Height Samples with Shortfall*. Münchner Wirtschaftswissenschaftliche Beiträge, 95/08.
- Komlos, J. (1989), *Nutrition and Economic Development in the Eighteenth Century Habsburg Monarchy: An Anthropometric History*. Princeton: Princeton University Press
- Komlos, J., u. J.H. Kim (1990), Estimating Trends in Historical Heights. *Historical Methods* 23/3, 116-120.
- Komlos, J. (1993a), Über die Bedeutung der anthropometrischen Geschichte. *Historical Social Research*, 18/3, 4-21.
- Komlos, J. (1993b), The Secular Trend in the Biological Standard of Living in the United Kingdom, 1730-1860. *Economic History Review*, XLVI/1, 115-144.
- Komlos, J. (1993c), Malthus, Boserup und wirtschaftliches Wachstum. Ein historischer Überblick. *Historical Social Research*, 18/4, 119-24.
- Maddala, G.S. (1983), *Limited-dependent and qualitative variables in econometrics*. Cambridge et al: Cambridge University Press.
- Malthus, T.R. (1798/1976), *An Essay on the Principle of Population*. 1. Aufl. 1798; Reprint New York: Norton.
- Trussell, J. & K. Wächter (1984), *Estimating the Covariates of Historical Heights*. NBER Working Paper Nr. 1455.
- Twarog, S.N. (1993), *Heights and Living Standards in Industrializing Germany: The Case of Württemberg*. UMI Dissertation Services Nr. 9325604.
- Wächter, K.W. (1981). Graphical Estimation of Military Heights. *Historical Methods*, 14/1, 31-42.
- Wächter, K.W. u. J. Tnjssell (1982), Estimating Historical Heights. *JASA*, 77/378, 279-293.