

IZA DP No. 14

**Memorandum: Erfolgsbedingungen empirischer  
Wirtschaftsforschung und empirisch gestuetzter  
wirtschafts- und sozialpolitischer Beratung**

Richard Hauser  
Gert G. Wagner  
Klaus F. Zimmermann

July 1998

# **MEMORANDUM: ERFOLGSBEDINGUNGEN EMPIRISCHER WIRTSCHAFTSFORSCHUNG UND EMPIRISCH GESTÜTZTER WIRTSCHAFTS- UND SOZIALPOLITISCHER BERATUNG**

**Richard Hauser  
Gert G. Wagner  
Klaus F. Zimmermann**

Discussion Paper No. 14  
July 1998

**Published in:  
Allgemeines Statistisches Archiv, 82 (1998), 369-379**

IZA

P.O. Box 7240  
D-53072 Bonn  
Germany

Tel: +49-228-3894-201  
Fax: +49-228-3894-210  
Email: [iza@iza.org](mailto:iza@iza.org)

This Discussion Paper is issued within the framework of IZA' s research area *General Labor Economics*. Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent, nonprofit limited liability company (Gesellschaft mit beschränkter Haftung) supported by the Deutsche Post AG. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public. The current research program deals with (1) mobility and flexibility of labor markets, (2) internationalization of labor markets and European integration, (3) the welfare state and labor markets, (4) labor markets in transition, (5) the future of work, and (6) general labor economics.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

## **ABSTRACT**

Memorandum:  
Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch  
gestützter wirtschafts- und sozialpolitischer Beratung

Based on the hypothesis that microdata are of crucial importance for internationally competitive economic research and scientific advice in economic and social policy, the authors develop various proposals for an improved data supply for empirical work in Germany. Present shortcomings are mainly caused by a lack of coordination between data collecting institutions and scientific users, extensive privacy protection provisions and restrictive information policy of government agencies. In order to overcome these problems, the authors ask for improved participation of scientific users in data collection; provision of recent techniques for data anonymization and decentralized use of confidential centralized data; liberalization of scientific data access; and creation of a documentation center for organization and standardization of data and data access. Finally, empirical economics, especially application of microdata and modern tools of data analysis, should be emphasized within the framework of university education.

JEL Classification: C81, C82

Keywords: microdata, data collection, data access, data organization

Klaus F. Zimmermann  
IZA  
P.O. Box 7240  
D - 53072 Bonn  
Tel: +49-228-3894-201  
Fax: +49-228-3894-210  
email: Zimmermann@iza.org

# Memorandum

## **Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung**

von

Richard Hauser

Johann Wolfgang Goethe-Universität, Frankfurt am Main

Gert G. Wagner

Deutsches Institut für Wirtschaftsforschung (DIW), Berlin, Europa Universität Viadrina, Frankfurt (Oder) und Centre for Economic Policy Research (CEPR), London

Klaus F. Zimmermann

Forschungsinstitut "Zukunft der Arbeit" (IZA), Bonn, Rheinische Friedrich-Wilhelms-Universität, Bonn, und Centre for Economic Policy Research (CEPR), London

In seinen "allgemeinen Gesichtspunkten" der Stellungnahme zu den Wirtschaftsforschungsinstituten der Blauen Liste in den alten Ländern hat der Wissenschaftsrat im Januar 1998 festgestellt, daß "im Zuge der Erarbeitung der vorliegenden Stellungnahmen [...] sich Hinweise darauf ergeben [haben], daß der Status der empirischen Wirtschaftsforschung in Deutschland - auch im internationalen Vergleich - Anlaß zu einer Reihe von Fragen gibt und daher eine breit angelegte Untersuchung zur Struktur, Organisation und Leistungsfähigkeit der empirischen wirtschaftswissenschaftlichen Forschung in Deutschland notwendig und förderlich ist. Der Wissenschaftsrat empfiehlt nachdrücklich, diese Untersuchung möglichst bald durchzuführen und Fragen der theoretischen Wirtschaftsforschung mit einzubeziehen, soweit dies für den Untersuchungszweck angebracht ist. Im

Anschluß daran sollte es zu weiteren Strukturempfehlungen kommen, die auch die Wirtschaftsforschungsinstitute der Blauen Liste einschließen".

Die Verfasser des vorliegenden Memorandums halten es für erforderlich, daß diese Analyse und Diskussion komplementär auch öffentlich durchgeführt wird. Zu einem zentralen Teilaspekt, der Datenproblematik, sollen die folgenden Überlegungen beitragen.

## **I. Theoretische Überlegungen**

Für die meisten wirtschafts- und sozialpolitischen sowie finanzwissenschaftlichen Fragestellungen reichen rein theoretische Modelle nicht aus, sondern politisch umsetzbare Ergebnisse bedürfen einer empirischen Fundierung. Umgekehrt sind empirische Messungen ohne ein theoretisches Gerüst im allgemeinen sinnlos. Empirische Wirtschaftsforschung und Wirtschaftstheorie (einschließlich der Finanzwissenschaft) sollten deswegen im Zusammenhang analysiert und im Hinblick auf ihre organisatorischen Strukturen betrachtet werden.

Empirische Wirtschaftsforschung im engeren Sinne zielt darauf ab,

- die relevanten Aspekte der wirtschaftlichen Aktivitäten mit Hilfe von Statistiken zu beschreiben und soweit wie möglich in quantitativer Form zu messen (deskriptive Statistik),
- aus statistischen Daten Hypothesen über Regelmäßigkeiten [Gesetze] abzuleiten (induktive Methode),
- Verhaltenshypothesen, die auf der Basis von theoretischen Modellen abgeleitet wurden, mit Hilfe von (vor allem) ökonometrischen Verfahren zu testen (deduktive Methode),
- internationale Vergleiche zu allen genannten Aspekte in deskriptiver Form anzustellen,
- die Unterschiede zwischen den Vergleichsländern als das Ergebnis "natürlicher" Sozialexperimente zu betrachten und auf dieser Basis die unterschiedlichen Auswirkungen verschiedener institutioneller Ausgestaltungen der Wirtschafts- und

- Sozialordnung für das Verhalten von Wirtschaftssubjekten herauszuarbeiten,
- Parameterwerte für einzelne Verhaltenshypothesen numerisch zu schätzen und auf dieser Basis quantifizierte Prognosen zu erarbeiten,
  - komplexe Modelle, die das wirtschaftliche Verhalten unterschiedlicher Akteursgruppen nachbilden und das Ergebnis des Zusammenwirkens dieser Gruppen nachzeichnen, zu erarbeiten,
  - mit Hilfe dieser numerisch spezifizierten Modelle Prognosen zu erstellen, und
  - mit Hilfe von Modellen Simulationsstudien über die Auswirkungen alternativer wirtschafts-, finanz- und sozialpolitischer Maßnahmen zu erstellen und damit empirisch gestützte Aussagen über die positiven und negativen Konsequenzen im Hinblick auf explizit gemachte Ziele abzuleiten.

Es ist offensichtlich, daß die Voraussetzung für eine erfolgreiche empirische Wirtschaftsforschung das Vorhandensein und die leichte Zugänglichkeit von möglichst vielfältigen und möglichst weit in die Vergangenheit zurückreichenden statistischen Daten ist.

Ebenso wie in den Naturwissenschaften ist auch bei der empirischen Wirtschaftsforschung eine möglichst breite nationale und internationale Konkurrenz zwischen Wissenschaftlern erforderlich, durch die sich die verlässlicheren deskriptiven Ergebnisse und die besseren Hypothesen durch gegenseitige Prüfung und Kritik allmählich durchsetzen. Aus diesen Gründen ist eine gleichrangige und gleichzeitige Zugänglichkeit von Datenbeständen für möglichst alle mit der jeweiligen Thematik befaßten Wissenschaftler erforderlich. Darüber hinaus sollten die Daten in einem engen Austauschprozeß zwischen analysierenden Forschern und Daten"produzenten" erstellt werden, um sicherzustellen, daß zu relevanten Fragen die angemessenen Daten mit methodisch adäquaten Verfahren erhoben werden.

Es geht uns im folgenden lediglich um die Erhebung und Verfügbarkeit der Daten für die wissenschaftliche Forschung und nicht-kommerzielle Nutzung.

## II. Organisatorische Ausgangslage

In Deutschland liegt die Verantwortung für die Datengenerierung und -nutzung bei der amtlichen Statistik, Administrationen, wissenschaftlichen Einrichtungen und privaten Unternehmen (insbesondere Marktforschungsinstituten). Die amtliche Statistik ist den Innenministerien bzw. Staatskanzleien nachgelagert bzw. sie ist Teil kommunaler und anderer Einrichtungen (z. B. Bundesbank, Bundesanstalt für Arbeit). Öffentlich finanzierte statistische Erhebungen gibt es außerhalb der Amtlichen Statistik im Bereich der Ressort- und Auftragsforschung und als Teil der Wissenschaftsförderung.

Die Erhebungsprogramme (und die zur Datenerhebung benutzten Methoden) werden von den analysierenden Forschern je nach Datenproduzent sehr unterschiedlich beeinflusst.

- Für die Daten der amtlichen Statistik gibt es in der Praxis fast keinen Einfluß der Wissenschaft auf die Erhebungsprogramme; diese Aussage gilt verstärkt für die Teile der amtlichen Statistik, die nicht in Statistischen Ämtern, sondern in Ministerien und anderen Einrichtungen geführt werden (z. B. Finanzministerium, Bundesanstalt für Arbeit, Bundesbank, Verband Deutscher Rentenversicherungsträger).
- Statistische Basisdaten, die mit öffentlichen Mitteln im Rahmen der Ressort- und Auftragsforschung erstellt werden (insbesondere auch in den Wirtschaftsforschungsinstituten), werden praktisch auch ohne Einfluß durch Wissenschaftler erstellt, die nicht an der Erhebung direkt beteiligt sind.
- Ein systematischer Einfluß der gesamten "Scientific Community" auf statistische Erhebungsprogramme gibt es nur dort, wo Daten durch wissenschaftliche Forschungsinstitutionen erhoben werden (z. B. DFG, VW Stiftung, teilweise auch BMBF).
- Auch auf die Erstellung vieler "abgeleiteter" Statistiken, wie insbesondere der Volkswirtschaftlichen Gesamtrechnung (VGR), gibt es keinen direkten Einfluß der Wissenschaft.
- Mangels professioneller Großerhebungen werden immer wieder Erhebungen von Studenten im Rahmen von „Lehrforschung“ durchgeführt (wobei das dilettantische

Verhalten von Studenten bei Interviews zur Gefährdung der Akzeptanz von statistischen Erhebungen bei Befragten beitragen kann).

Auch die Zugänglichkeit dieser Datenbestände für die nicht-kommerzielle, empirische Wirtschaftsforschung an Universitäten und gemeinnützigen Forschungsinstituten unterscheidet sich deutlich. Für diese Unterschiede sind maßgeblich: (1) die Art der Finanzierung, (2) Datenschutzbestimmungen, (3) Verhaltensweisen der Datenproduzenten, und schließlich (4) die Verfahren der Datensammlung, Datenaufbereitung, Datendokumentation und Datenübermittlung, die angesichts der tiefgreifenden Änderungen bei den elektronischen Netzwerken und der Verfügbarkeit von bisher nicht vorhandenen Rechen- und Speicherkapazitäten am Arbeitsplatz des Wissenschaftlers einem tiefgreifenden Wandel unterliegen.

- Weitgehend unregelt ist der Zugang zu statistischen Daten, die vom Staat - außerhalb der eigentlichen amtlichen Statistik - in Ministerien und anderen Einrichtungen produziert werden.
- Der Zugang zu Daten der amtlichen Statistik ist weitgehend geregelt, allerdings gibt es zwei grundsätzliche Probleme: (1) Mikrodaten sind zum größten Teil nicht oder nur schwer zugänglich; dies gilt insbesondere für die nur schwer anonymisierbaren Betriebs- und Unternehmensdaten der Wirtschaftsstatistik. Mit den vom BMBF finanzierten "Scientific Use Files" für einige Stichproben der Bevölkerungsstatistik gibt es allerdings eine neue Entwicklung für den regelhaften Zugang zu anonymisierten Mikrodaten; (2) aggregierte Daten, z. B. der VGR, sind regelhaft zugänglich, insbesondere sind auch die Veröffentlichungszeitpunkte allgemein bekannt, jedoch ist für Außenstehende nicht klar, ob und wann Statistische Ämter vorläufige Daten und Sonderauswertungen zur Verfügung stellen. Manchmal führt dies dazu, daß Wissenschaftler den Eindruck gewinnen, daß Daten willkürlich herausgegeben oder zurückgehalten werden.
- Weitgehend unregelt ist auch der Zugang zu Daten, die mit öffentlichen Mitteln außerhalb der amtlichen Statistik erhoben werden. Zwar werden viele Erhebungen dem "Zentralarchiv für empirische Sozialforschung" in Köln übergeben, oder die Daten werden von den Produzenten direkt zugänglich gemacht, aber es ist oft zufällig, ob und wie dies geschieht. Nur wenige Wissenschaftsförderungsorganisationen knüpfen die Finanzierung einer Erhebung



an die explizite Verpflichtung, diese rasch für Re-Analysen zugänglich zu machen. Vorbildlich ist hier beispielsweise die DFG.

Zu obigen Punkten nimmt das Memorandum im folgenden Stellung, um die Verantwortlichen zu einer den neuen wissenschaftlichen Entwicklungen entsprechenden Gestaltung der statistischen Infrastruktur aufzurufen und damit die Erfolgsbedingungen für die empirische Wirtschaftsforschung und empirisch gestützte Politikberatung zu verbessern.

### **III. Datentypen und Datenschutz**

Vorweg ist festzustellen, daß sich die empirische Wirtschafts- und Sozialforschung in dem oben beschriebenen Sinn nicht für identifizierbare Einzelfälle, sondern lediglich für anonyme Untersuchungseinheiten, für nach bestimmten Kriterien zusammengefaßte Gruppen von Untersuchungseinheiten und für statistische Regelmäßigkeiten interessiert. Individualisierte Fallstudien liegen außerhalb ihres Untersuchungsbereichs; dementsprechend sind die Probleme des Datenschutzes und der Datenzugänglichkeit auch nur für diese begrenzte statistische Nutzung zu klären.

Aus der Sicht des Datenschutzes lassen sich die für die empirische Wirtschaftsforschung erforderlichen Daten in zwei Klassen einteilen:

1. In Tabellenform oder summarisch zusammengefaßte Daten, wie sie etwa in der Volkswirtschaftlichen Gesamtrechnung, in Finanzstatistiken der Deutschen Bundesbank oder in Bevölkerungsstatistiken des Statistischen Bundesamtes vorliegen.
2. Einzeldaten von Personen, Haushalten, Unternehmen, Betrieben, Organisationen und staatlichen Organen.

Die erste Kategorie bietet keine datenschutzrechtlichen Probleme, sofern in den

Tabellen eine genügende Anzahl von Untersuchungseinheiten zusammengefaßt sind und nur Durchschnitte oder Summen ausgewiesen werden. Für diese Daten bestehen lediglich die Probleme

- einer nutzerfreundlichen Information über die verfügbaren aggregierten Daten und Zeitreihen,
- der Standardisierung der Konzepte, der korrekten Erfassung und der Vergleichbarkeit in der zeitlichen, regionalen und internationalen Dimension,
- der verlässlichen (regelgebundenen) und schnellen Zugänglichkeit am Arbeitsplatz des Wissenschaftlers (Dazu gehören auch Regeln, wie und wann vorläufige Daten herauszugeben sind; es stehen inzwischen auch Computeralgorithmen zur Verfügung [Statistical Disclosure Control], die das automatische, d.h. rasche Prüfen und ggf. eine geeignete Aggregation dünn besetzter Zellen vornehmen).

Diese Punkte werden im Memorandum unten nochmals aufgegriffen.

Die zweite Kategorie der Einzeldaten ist in individualisierter Form (mit Identifikationsmerkmalen wie Name, Adresse, Telefonnummer) aus Datenschutzgründen grundsätzlich nicht zugänglich. Im Hinblick auf den Charakter der Information lassen sich folgende Typen von Datensätzen unterscheiden:

- Einmalige Querschnittserhebungen;
- Zeitreihen von Querschnittserhebungen, wobei im Prinzip jeweils unterschiedliche Untersuchungseinheiten die Informationen liefern;
- Wiederholungsbefragungen derselben Untersuchungseinheiten (Panels), die über einen kürzeren oder längeren Zeitraum fortgeführt werden.

Wenngleich Querschnittserhebungen und Zeitreihen von gleichartigen Querschnittserhebungen seit vielen Jahrzehnten eine unverzichtbare Basis für die empirische Wirtschaftsforschung darstellen, haben sich in den vergangenen beiden Dekaden doch Panelerhebungen von Mikrodaten als besonders fruchtbare Informationsquellen für viele Fragestellungen erwiesen, die mit Querschnittsdaten nicht geklärt werden konnten. Gerade auf diesem Gebiet vollzieht sich zur Zeit ein besonders schneller wissenschaftlicher Fortschritt, da hiermit die Reaktionen von Untersuchungseinheiten auf Änderungen der Einflußgrößen wesentlich besser analysiert werden können.

Auch mit Daten, die im Verwaltungsprozeß anfallen und die überdies mit Umfragedaten kombiniert werden können, sind bereits vielversprechende Untersuchungen angestellt worden. Auf diesem Feld liegen ebenfalls große, bisher nicht ausgeschöpfte, Möglichkeiten für die empirische Wirtschaftsforschung.

Es sind in den in den vergangenen beiden Jahrzehnten verschiedene Verfahren entwickelt worden, jene Informationen, die die empirische Wirtschaftsforschung braucht, zur Verfügung zu stellen, ohne daß eine Identifizierung des Merkmalsträgers erfolgen kann. Diese Verfahren sind im Statistikgesetz abgesichert. Dabei handelt es sich um

- Verfahren der absoluten Anonymisierung und um
- Verfahren der faktischen Anonymisierung.

Auf dieser Basis konnten inzwischen große amtliche Stichproben von Personen und Haushalten (Einkommens- und Verbrauchsstichproben, Mikrozensus) an wissenschaftliche Nutzer weitergegeben werden. In diesem Zusammenhang bestehen aber noch die Probleme

- der Kontinuität der Weitergabe,
- der Beschleunigung des Weitergabeverfahrens,
- die Standardisierung der Datenweitergabe, um die Installationskosten der interessierten Wissenschaftler möglichst gering zu halten,
- der Kostenübernahme für die faktische Anonymisierung,
- einer verbesserten Dokumentation der Erhebungs- und Aufbereitungsarbeiten durch das Statistische Bundesamt,
- der Beratung bei Datenproblemen über die Dokumentation hinaus und einer Rückkoppelung zur Verbesserung der Erhebungsinstrumente und der Datenaufbereitung,
- der Kooperation zwischen Statistischem Bundesamt und der Wissenschaft bei der Datenprüfung und -bereinigung,
- der Regelung der Einflußnahme auf die Erhebungsprogramme durch die Wissenschaft, um die theoretisch relevanten Phänomene und Kausalfaktoren besser erfassen zu können, und schließlich

– der Ausdehnung der Weitergabeverfahren auf andere Träger von Datenerhebungen und Besitzer von mit öffentlichen Mitteln beschafften Einzeldatenbeständen.

Es gibt auch Einzeldatensätze für Untersuchungseinheiten, insbesondere Unternehmen, Organisationen und staatliche Stellen, die sich nicht faktisch anonymisieren lassen, weil die Anzahl der jeweiligen Merkmalsträger so gering ist, daß sie bereits mit wenigen Zusatzinformationen identifiziert werden können oder daß zumindest einige herausragende Merkmalsträger identifiziert werden können. Beispiele wären Arbeitsstättenzählungen, Betriebspanels, u. ä. Um diese Einzeldaten für die empirische Wirtschaftsforschung zugänglich zu machen, bedarf es anderer Verfahren, wie beispielsweise die sanktionsbewehrte Verpflichtung zur vertraulichen Behandlung von Einzeldaten durch den Forscher oder die temporäre Beschäftigung bei einem Datenbesitzer mit entsprechender Verpflichtung zur Wahrung der Vertraulichkeit.

#### **IV. Thesen zur Finanzierung und Organisation**

1. Daten können mehrfach eingesetzt werden. Ihre Produktionskosten lohnen sich häufig nur, wenn sie in vollem Umfang genutzt werden. Ein freier Markt wird dieses Produkt nur in unzureichendem Umfang anbieten. Statistische Daten sind quasi "öffentliche Güter", die öffentlich finanziert werden müssen - was einen regulierten Wettbewerb bei der Produktion dieser Daten freilich keineswegs ausschließt. Eigentümer ist die Gesellschaft, die die Daten nur aus organisatorischen Gründen in den Besitz unterschiedlicher Institutionen gibt.
2. Die Daten der amtlichen Statistik sind offensichtlich alle öffentliche Güter in diesem Sinne.
3. Die Daten der öffentlichen Verwaltung, die für die administrative Vorgänge erhoben werden oder dabei anfallen (Steuerdaten, Daten im Bereich des Arbeitsmarktes und des sozialen Sicherungssystems), sind ebenfalls öffentliche Güter; diese Daten sollten unter Mitarbeit von unabhängigen Wissenschaftlern soweit wie dies wissenschaftlich sinnvoll und datenschutzrechtlich möglich ist, als anonymisierte Datensätze aufbereitet und ebenfalls soweit wie möglich in geeigneter Form der Wissenschaft zur Verfügung gestellt werden.

4. Sofern die Datenproduktion direkt oder indirekt über Forschungsförderungsinstitutionen mit öffentlichen Mitteln erfolgte, sollten die Daten allen Wissenschaftlern auf Anforderung gegen einen geringen Kostenbeitrag zur Verfügung gestellt werden.
5. Für internationale Datenbasen sind geeignete Finanzierungsformen zu finden.
6. In Deutschland ist zu prüfen, ob die Anbindung der Amtlichen Statistik an die Innenministerien bzw. Staatskanzleien, sowie die "Auslagerung" von Amtlicher Statistik in Ministerien, die Bundesbank, die Bundesanstalt für Arbeit und in einige Verbände optimal ist. Generell sind eine Trennung der Datenproduktion von politischen Bürokratien in unabhängige Institutionen mit wissenschaftlichen Beiräten und eine Dezentralisierung zweckmäßig.

#### **V. Thesen zur Festlegung von Erhebungsprogrammen**

1. Die Erhebungsprogramme für öffentlich relevante statistische Daten müssen von den Vertretern der Öffentlichkeit zusammen mit unabhängigen Fachwissenschaftlern (für Erhebungsmethoden und Analysen) entwickelt werden.
2. Es ist nicht nur für die Erhebungen der Amtlichen Statistik im engeren Sinne notwendig, daß Erhebungsprogramme von der Öffentlichkeit und den Fachwissenschaften beeinflußt werden können, sondern auch für Prozeßstatistiken und für Erhebungen, die im Rahmen der Ressortforschung, in öffentlich finanzierten Instituten und in Form von öffentlich finanzierten Aufträgen erhoben werden.
3. Fachbeiräte sind nützliche Instrumente, um die Mitwirkung der Öffentlichkeit und der Wissenschaft bei der Festlegung von Erhebungsprogrammen zu sichern.

#### **VI. Thesen zur Beherrschung von Datenschutzproblemen**

1. Die Möglichkeiten der "faktischen Anonymisierung" von statistischen Mikrodaten müssen konsequent genutzt werden. Sub-Sampling (also die Weitergabe einer Unterstichprobe der ursprünglich erhobenen Daten) ist eine wirksame Strategie, um Datenschutz einzuhalten, ohne daß dadurch Ergebnisse verzerrt werden. Hinzu kommen Möglichkeiten, Variablen verzerrungsfrei zu vergrößern und die

Möglichkeit, besonders sensitive Variablen nicht zugänglich zu machen (z. B. feinräumige Regionalinformationen). Letztgenannte Methoden verzerren zwar keine Ergebnisse, machen jedoch oft aussagekräftige Untersuchungen unmöglich.

2. Insoweit die faktische Anonymität von Mikrodaten nur herstellbar ist, indem wesentliche Variablen vergrößert oder nicht weitergegeben werden, müssen Datenproduzenten, die mit öffentlichen Mitteln arbeiten, verpflichtet werden, die Arbeit mit sensitiven Daten innerhalb ihrer Institutionen zu ermöglichen; dazu gehören ggf. auch "Filialen" dieser Einrichtungen, die in der Praxis den Zugang wesentlich erleichtern können. Insoweit aus Datenschutzgründen nur Mitarbeiter einer Einrichtung mit bestimmten Daten arbeiten dürfen, müssen entsprechende Anstellungsverträge auch für „One-Dollar-Persons“ ermöglicht werden (Fellowship-Programme nach US-amerikanischem Muster).
3. Darüber hinaus ist zu prüfen, ob die Ökonometrie- und Statistik-Ausbildung in den Universitäten derart professionalisiert werden kann, daß für Wissenschaftler ein bindender Ethik-Code zur Wahrung des Datengeheimnisses etabliert werden kann (ähnlich wie für Ärzte, Rechtsanwälte usw.), der bei Nicht-Einhaltung zum Entzug der "Lizenz" für empirisches Arbeiten führt. Ein derartiger Wissenschaftler-Ethik-Code muß mit entsprechenden rechtlichen Rahmenbedingungen verbunden werden, die sicherstellen, daß für übergeordnete staatliche Zwecke das „Statistik-Geheimnis“ nicht gebrochen werden kann. Bei einer derartigen Entwicklung muß beachtet werden, daß sie nicht zu einer Verknappung der entsprechend lizenzierten Wissenschaftlern führt, indem Marktzutrittsbarrieren errichtet werden, um der Zukunft allokativ ungerechtfertigte Einkommensvorteile zu sichern.

## VII. Thesen zum Ausbau der Infrastruktur

### • Thesen zur ersten Datenkategorie (aggregierte Daten)

1. Sichergestellt werden muß eine *regelgebundene* Verfügbarkeit aggregierter Ergebnisse. Die Regelgebundenheit gilt auch für vorläufige Daten und Sonderauswertungen. Zumindest die endgültigen Daten sollten on-line zur Verfügung gestellt werden, und zwar sowohl nationale Daten als auch die von internationalen Organisationen gesammelten und aufbereiteten Daten

2. Gegebenenfalls ist die Einrichtung einer deutschen Clearingstelle notwendig, die alle Daten beschafft und den Bestand pflegt, sofern das Statistische Bundesamt diese Aufgabe nicht übernimmt.
3. Um Zugangshürden abzubauen (insb. für Universitätsinstitute, die nicht permanent mit bestimmten Daten arbeiten) ist für die Wissenschaft eine Standardisierung des Datenzugangs notwendig, verbunden mit geringen Gebühren oder Kostenbefreiung.

- **Thesen zur zweiten Datenkategorie (Mikrodaten)**

4. In Zusammenarbeit mit den Forschungsförderungseinrichtungen (insb. der DFG) sollten zügig Fellowship-Programme eingerichtet werden, die das Arbeiten mit nicht oder nur schwer anonymisierbaren Mikrodaten bei den datenerhebenden Einrichtungen, die die datenschutzrechtliche Verantwortung tragen.
5. Um Zugangshürden abzubauen (insb. für Universitätsinstitute, die nicht permanent mit bestimmten Daten arbeiten) ist für die Wissenschaft eine Standardisierung des Zugangs zu Mikrodaten (amtlicher und nicht-amtlicher Art) mit geringen Kosten oder Kostenbefreiung notwendig.
6. Ein weiterer Schritt wäre die Einrichtung einer Clearingstelle, die in regelhafter Weise Informationen über sämtliche Einzeldatenbestände und die jeweiligen Dokumentationen bereithält und die on-line konsultiert werden kann; über die Weiterentwicklung dieser Clearingstelle zu einer Datenbeschaffungsagentur für die empirische Wirtschaftsforschung sollte nachgedacht werden.
7. Es muß geprüft werden, wie getrennt erstellte statistische Datenbasen in datenschutzrechtlich einwandfreier Weise zusammengeführt werden können ("Link", z.B. von Firmen- und Bevölkerungsinformationen). Unter Umständen können Clearing-Stellen auch für diesen Zweck nützlich sein.
8. Geprüft werden sollte unter datenschutzrechtlichen Gesichtspunkten die Möglichkeit des "Fernrechnens" für nicht anonymisierbare Datenbestände mit entsprechenden Sicherungsmaßnahmen gegen die Identifizierung einzelner Merkmalsträger. „Fernrechnen“ bedeutet hier Übermittlung von Auswertungsprogrammen on-line an den Datenbesitzer und Rückgabe von geschätzten Parameterwerten sowie tabellierten Ergebnissen (unter Berücksichtigung von datenschutzrechtlich gebotenen Mindestfallzahlen).

## **VIII. Thesen zum Ausbau der empirischen Wirtschaftsforschung an den Universitäten**

1. Stärkung der empirischen Wirtschaftsforschung in den Fächern Wirtschaftsund Sozialpolitik/Finanzwissenschaft und Statistik/Ökonometrie.
2. Einbau von Datenanalyse, Datenkritik und Datenschutz in die mittlere Ausbildungsphase von Ökonomiestudenten.
3. Verbesserte Ausbildung an Statistikpaketen.
4. Arbeit mit echten Daten bereits im Studium, evtl. in Zusammenarbeit mit Datenproduzenten, mit Hilfe von kleinen, hochanonymisierten Substichproben; Verzicht auf dilettantische Primärerhebungen durch Studenten, die die Akzeptanz von Erhebungen bei Befragten gefährden.
5. Verbesserung des Datenservice an den Universitäten für die Studenten und ausreichende Verfügbarkeit der jeweiligen Statistikpakete und Computerkapazitäten.
6. Zugang zu Clearingstellen und "Fernrechenmöglichkeiten" für Studenten.

## **IX. Aufforderung zur Diskussion**

Dieses Memorandum soll zur Diskussion anregen. Diese kann in den einschlägigen Fachzeitschriften und auf speziellen Workshops geführt werden. Die Verfasser sind auch für direkte Stellungnahmen dankbar. Ergebnisse dieser Diskussion könnten dann auch für die Arbeit der Gremien des Wissenschaftsrates Gestaltungshilfen liefern.

**Richard Hauser, Gert G. Wagner und Klaus F. Zimmermann**

**Frankfurt am Main, Berlin und Bonn**

**Juni 1998**



### **Kontaktadressen:**

- Prof. Dr. Richard Hauser  
Johann Wolfgang von Goethe-Universität  
Fachbereich Wirtschaftswissenschaften  
Senckenberganlage 31  
60325 Frankfurt am Main  
e-mail: R.Hauser@em.uni-frankfurt.de  
Tel.: 069/7982-2546  
Fax: 069/7982-8287
  
- Prof. Dr. Gert G. Wagner  
Deutsches Institut für Wirtschaftsforschung  
Königin-Luise-Straße 5  
14195 Berlin  
e-mail: gwagner@diw.de  
Tel.: 030/89789-290  
Fax: 030/89789-109
  
- Prof. Dr. Klaus F. Zimmermann  
Forschungsinstitut zur Zukunft der Arbeit (IZA)  
Postfach 7240  
53072 Bonn  
Tel.: 0228/3894-200  
Fax: 0228/3894-210