

Applied Economics Quarterly Supplement

2004

Beihefte der
Konjunkturpolitik

Issue 55

The European Reform Logjam
and the Economics of Reform

Edited by
Thomas Straubhaar
and Rainer Winkelmann



Duncker & Humblot · Berlin

DOI <https://doi.org/10.3790/978-3-428-51659-9>

Generated for Hochschule für angewandtes Management GmbH at 88.198.162.162 on 2025-06-09 07:54:24

FOR PRIVATE USE ONLY | AUSSCHLIESSLICH ZUM PRIVATEN GEBRAUCH

The European Reform Logjam and the Economics of Reform

Applied Economics Quarterly Supplement

Beihefte der Konjunkturpolitik

Issue 55

The European Reform Logjam and the Economics of Reform

Edited by

Thomas Straubhaar
and Rainer Winkelmann



Duncker & Humblot · Berlin

DOI <https://doi.org/10.3790/978-3-428-51659-9>

Generated for Hochschule für angewandtes Management GmbH at 88.198.162.162 on 2025-06-09 07:54:24
FOR PRIVATE USE ONLY | AUSSCHLIESSLICH ZUM PRIVATEN GEBRAUCH

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in
the Deutsche Nationalbibliografie; detailed bibliographic data
is available in the Internet at <<http://dnb.ddb.de>>.

Technical editor: Deborah Anne Bowen

All rights reserved

© 2004 Duncker & Humblot GmbH, Berlin

Typesetting and printing:

Berliner Buchdruckerei Union GmbH, Berlin

Printed in Germany

ISSN 1612-2127

ISBN 3-428-11659-3

Gedruckt auf alterungsbeständigem (säurefreiem) Papier
entsprechend ISO 9706 ☺

Internet: <http://www.duncker-humblot.de>

DOI <https://doi.org/10.3790/978-3-428-51659-9>

Editorial

This supplement to Applied Economics Quarterly reports on the 67th Annual Meeting of the Association of German Economics Research Institutes (ARGE) in association with other European research institutes. The Annual Meeting took place in Berlin on April 22 – 23, 2004. The topic was

“The European Reform Logjam and the Economics of Reform”

Daniel Gros (CEPS), Carsten Hefeker (HWWA) and Yvonne Heiniger (HWWA) were responsible for the conceptual preparation of the conference. The opening address was given by Daniel Gros (CEPS). Subsequent sessions were organized in the form of presentations, each followed by a discussant’s statement. The following contributed to the conference: Norbert Berthold (Bayerische Julius-Maximilians Universität Würzburg), Charles B. Blankart (Humboldt-Universität, Berlin), Mathias Erlei (Technische Universität Clausthal), Michael Fertig (RWI, Essen), Daniel Gros (Centre for European Policy Studies, Brussels), Peter Haug, (IWH, Halle), Friedrich Heinemann (ZEW, Mannheim), Jochen Kluve (RWI, Essen), Rigmor Osterkamp (ifo, Munich), Martin T.W. Rosenfeld (IWH, Halle), Ronnie Schöb (Otto-von-Guericke-Universität, Magdeburg), Viktor Steiner (DIW Berlin), Thomas Straubhaar (HWWA, Hamburg), Peter Zweifel (University of Zurich).

Next year’s Annual Meeting is scheduled for April 14, 2005 in Berlin and will deal with issues of locational competition.

June 2004

*Thomas Straubhaar
Rainer Winkelmann*

Contents

Friedrich Heinemann

Explaining Reform Deadlocks	9
Comment: <i>Mathias Erlei</i>	27

Peter Haug and Martin T. W. Rosenfeld

The Reform of Local Public Services of General Interest in Europe	31
Comment: <i>Charles B. Blankart</i>	51

Viktor Steiner

Social Welfare Reform and the Low-Wage Labor Market in Germany: What Works and What Doesn't?	57
Comment: <i>Norbert Berthold</i>	79

Michael Fertig and Jochen Kluge

A Conceptual Framework for the Evaluation of Comprehensive Labor Market Policy Reforms in Germany	83
Comment: <i>Ronnie Schöb</i>	113

Rigmar Osterkamp

Health-Care Efficiency in OECD Countries	117
Comment: <i>Peter Zweifel</i>	143

Participants	151
--------------------	-----

Explaining Reform Deadlocks

By Friedrich Heinemann*

Abstract

Countries are often slow to adjust their economic structures to new necessities although this reform reluctance is costly in terms of growth and employment. This paper analyses the relevant factors that block or foster economic reforms. Theoretical considerations show that there are at least three classes of potentially relevant factors: the objective need for reforms, political-economic issues, and factors associated with limited rationality or rational ignorance. In the empirical analysis, a reform event is quantified as a significant change of the Economic Freedom of the World indicator within a five-year-period. This makes it possible to run probit estimations for a large country panel starting in the 1970s where the probability of reform is explained by a number of proxies covering all three classes of potential factors. The results suggest that the initial extent of economic freedom and growth performance are empirically relevant factors. Furthermore, countries with an ageing population appear to behave in a less reform-friendly manner.

JEL-Classification: E63, H00

Keywords: economic policy reforms, economic freedom, reform resistance, limited rationality

1. Introduction

Why are countries frequently so slow to adjust their economic structures to new necessities although this reform reluctance leads to substantial losses in growth and employment? It is currently of crucial importance to find a convincing answer to this question.

Particularly, in big EU countries like Germany, France and Italy there appear to be huge lags between identification of economic problems and the implementation of political remedies. Examples concern highly regulated labour markets, an excessive tax burden and the exploding costs of the social security systems not adjusted

* Centre for European Economic Research (ZEW), Mannheim Friedrich Heinemann, P.O. Box 10 34 43, 68034 Mannheim, Germany. Tel: + 49-621-1235-149; Fax: + 49-621-1235-223; Email: heinemann@zew.de

The author gratefully acknowledges helpful comments from Mathias Erlei and Rainer Winkelmann.

to the needs of an ageing population. That these factors are at the root of growth problems in many countries is largely uncontroversial in spite of all debates in detail. Given this principal consensus it is hard to explain why a number of countries are so hesitant to change structures even if the costs of the reform deadlock have become clearly visible in the growth and labour market statistics.

This phenomenon is a challenge not only for economic policy but also for economic theory: It is no easy task to reconcile this self-destructive behaviour of whole societies with the central economic assumption of rationality. Of course, political-economic theory possesses concepts and models to explain under certain assumptions “stupid” social decision-making in spite of individual rationality. Explanations surveyed in brief below range from Down’s notion of “rational ignorance” to Alesina’s and Drazen’s “war of attrition”. However, this class of elegant explanations loses its convincing power, the more visible, relevant, and certain the costs of the reform deadlock become for an overwhelming majority of the population.

In some instances, the refusal to acknowledge basic economic realities appears so widespread among populations that it is not easy to stick faithfully to the assumption that voters are rational and optimising individuals. The debate on the pension age is one example: the dramatic increase in life expectancy, low birth rates, and the resulting ageing of the population make raising the pension age an unavoidable measure to limit the exploding dependency ratios in the coming decades. Nevertheless overwhelming majorities of all age groups in the population are opposed to this measure according to recent opinion polls, for example, in Germany.

In the field of financial economics, behavioural economists have made use of non-rational psychological patterns in human behaviour to explain financial market phenomena. It is amazing that the same instruments have rarely been applied to the analysis of economic policy. Of course, it is a challenge for economists to debate the rationality assumption. However, given the fact that irrationalities are observable in financial market decision-making, there is no reason to exclude the possibility of irrational behaviour a priori as an explanation for economic policy failures.

Given these profound intellectual and political problems, our understanding of the true reasons behind today’s reform deadlocks is absolutely dissatisfactory. These research deficiencies are particularly marked in the empirical field as Rodrik (1996, 25) complains: “We now need theoretically informed case studies – or more formal tests – that attempt to discriminate among these alternative stories”. This study follows this recommendation, not in regard to case study techniques, but by focusing on more formal testing based on a large country panel. With the exception of Ptilik and Wirth (2003), to our knowledge there are no comparable attempts in the literature. Close to our approach, these authors endogenise economic reforms but focus on the hypothesis that crisis is a precondition for reforms

without showing interest into a wide spectrum of rational and irrational types of reform obstacles.

This paper's empirical approach is the following: based on a panel of 123 countries, we analyse which factors can explain that some countries experience significant reform eras and others do not. The occurrence of reforms is measured on the basis of Fraser Institute's Economic Freedom of the World (EFW) index. Both the aggregate index and the sub-components are analysed so that differentiated insights for different fields of economic policy (size of government, property rights, monetary regime, external freedom and regulation) can be derived. Among the potential explaining factors, a distinction is made between three classes of variables: first, proxies for the objective necessity of reform; second, proxies for political-economic reform obstacles; and third, proxies that hint at the relevance of limited information or limited rationality.

The structure is as follows. In Section 2, we give a brief survey on the classes of reform obstacles discussed in the literature. Section 3 describes the variables and gives a first descriptive view of the data. Section 4 presents the econometric analysis. In the final section, we draw conclusions.

2. Reform Obstacles

Explanations of why societies do not accept economic reforms that are beneficial for a large majority of citizens can be divided into those that stick to the rationality assumption and those that do not.

Explanations based on rational reform resistance

Examples of rational explanations are the following (for a survey see Rodrik 1996): reforms might be beneficial to the population as a whole in the sense of higher growth rates and welfare, but nevertheless, distributive consequences are unavoidable. There are always groups that suffer from adjusting economic structures. The standard interest group explanation related to these distributive aspects is that minorities that suffer from reforms are well organised and powerful enough to block any change hurting them. The distributive consequences might interact with discount rates differing between groups in society. If certain reform projects have a J-curve effect – short-run losses followed by long-term gains – groups with a short time horizon and high discount rates are likely to oppose these projects. These inter-temporal aspects are likely to be of particular relevance in ageing societies where the growing higher age segments of the population apply a heavy discounting.

Uncertainty about a reform's outcome – possibly coupled with risk-aversion – can pose a further obstacle to reforms: Fernandez and Rodrik (1991) show that even with risk-neutral voters, uncertainty about the winners' identity can lead to

the democratic rejection of reforms that are beneficial for a majority. Models that focus on incomplete information and coordination problems show that reforms might be blocked even though this implies that the politically powerful groups hurts themselves. Alesina's and Drazen's "war of attrition" (1991) is the classic example in this field: two groups are uncertain about the opponent's cost of stabilisation and both have an incentive to delay reforms. These examples of rational reform resistance can be relevant under both democratic regimes and dictatorships.

Rational ignorance (Downs 1957) is a further variant of rational explanations for detrimental institutional stickiness. According to the Downsian notion, voters' individual information optimisation results in socially inefficient information activities. The consequence is that voters are badly informed about the consequences of reform options and about the costs of the status quo.

While these rational explanations are elegant and without doubt hint at important real dimensions of the reform problem, they do not appear to be fully satisfactory. Indeed, it is not clear at all why reforms should necessarily increase uncertainty about the economic future of individuals or certain interest groups. This is particularly questionable if the status quo is not sustainable. With an ageing population, sticking to the status quo, for example in the field of pay-as-you-go pension systems, creates significant uncertainty about future pension payments. Here, reforms that create a more sustainable system are likely to be conducive to the credibility of the promised pension. In such cases, reforms can help limit uncertainty.

Explanations based on the distributive consequences of reforms are not always convincing either. Countries that currently find themselves in acute reform deadlocks suffer from significant growth and welfare losses that hurt both a large majority of the population and many politically powerful interest groups. Here it is hard to see which groups are reform losers and at the same time powerful enough to block reforms against all those who would benefit. Interest-group-related explanations also face difficulties in explaining why reform winners cannot compensate the politically influential minorities that oppose reforms. The higher the costs to the status quo, the larger the incentive is for the winners to pay such compensation.

Finally it is doubtful whether rational ignorance can really establish a systematic bias against reforms. A low level of information means that people have wrong expectations about the consequences of reforms, yet there is no obvious argument why this should necessarily imply any bias against reforms. Without further arguments, it is equally likely that poor information is associated with an overly euphoric assessment of reforms (Wittman 1995). In any case, the rational ignorance argument weakens the more visible the cost of institutional stickiness becomes.

Explanations based on limited rationality

Given the above-mentioned difficulties of rational explanations, a fundamentally different class of approaches gives up the assumption of full rationality. Pro-

ponents (Caplan 2001) state their analytical starting point in the following way: “[Voters] oppose better policies not from complicated strategic calculations, but because they don’t understand what works. . . . What dominates political debate and public opinion is not subtle strategizing, but elementary economic misconception.”

Behavioural patterns that constitute deviations from rationally optimising behaviour have frequently been applied in financial market economics (see for example Daniel, Hirshleifer and Teoh 2002). It is amazing that the success of the “behavioural finance” field is not yet paralleled by a field like “behavioural economic policy”. Compared to financial markets, human instincts and psychologically rooted irrationalities should have a deeper impact in economic policy phenomena due to the lack of market forces that punish irrationalities on an individual basis. Compared to an instinct-driven voter, a voter who practices rational optimisation at the ballot box does not benefit from better politics. From an individual perspective, irrational voting is extremely cheap; the marginal costs are basically zero. The case is different for private economic decisions, for example on investment, the labour market, and education. Choosing an education without job and income chances, for example, is punished on an individual basis and thus demands a high individual price. Choosing a political program at the ballot box that is doomed to economic failure is individually costless.

Heinemann (2001) argues that a number of empirically proven psychological anomalies have a substantial explanatory potential in the context of reform resistance. Examples are phenomena such as the “status quo bias”, the “endowment effect” and “loss aversion”. The status quo bias describes a situation where people have a preference for one option among many others only because this option happens to be the status quo (Samuelson and Zeckhauser 1988). If a change occurs, this specific option loses attraction immediately. The endowment effect stands for preferences that depend on whether a certain good is possessed or not. Experiments show that the willingness to pay for acquiring good x is significantly smaller than the price the same person would accept for giving up good x (Kahneman et al. 1991). Loss aversion, finally, denotes the fact that the absolute change in utility associated with a loss is larger than the absolute change in utility associated with a gain (Tversky and Kahnemann 1991). With loss aversion, the utility function is non-continuous in the reference point, which tends to be the status quo. All these anomalies work against reforms that change the status quo. Sometimes these anomalies also underline arguments familiar from rational explanations: Loss aversion, for example, might explain why the reform resistance of reform losers is more intense than the reform support from reform winners.

Empirically it is hard to disentangle rational ignorance from rational irrationality since both phenomena are deeply interrelated within the human mind. However, both concepts hint at the same proxies that can be included in empirical studies, for example variables on the state of a population’s education.

This short overview demonstrates that there is a whole universe of factors that might potentially be relevant for the explanation of reform speed. There is no single empirical approach that would be able to cover all these dimensions. Hence, in the following, we apply a less ambitious approach. We look for empirical evidence of the relevance of three classes of factors:

- First, factors that represent the objective need for reforms and thus correspond to rational explanations in the sense that increasing costs of institutional stickiness should help overcome reform resistance.
- Second, political-economic factors that explain the relative power of reform-resistant groups and thus support rational explanations linked to the distributional consequences of reforms.
- Third, factors that correspond to rational ignorance and/or limited rationality, two classes of explanations that can hardly be disentangled on the basis of the macro-approach applied here.

3. Variables and Descriptive Analysis

Any attempt to study the economics of reform processes using statistical and econometric tools faces a serious data problem. Adjusting an economy's structure to new requirements is a complex process whose quantification is no easy task. The problems are manifold: It must be decided what kinds of institutional change are relevant and appropriate, and what kinds are not. Furthermore, changes must be translated into numbers that offer a starting point for the empirical assessment. Finally, the quantification should be available for different countries and different time periods in order to make inter-country comparisons and analyses along the time axis.

To cope with these problems, this study bases its quantification of reform processes on the Economic Freedom of the World (EFW) index developed by the Fraser Institute (Gwartney and Lawson 2003a,b). The content of the index has developed over time. It is available since 1970 in intervals of five years and covers 123 countries. In its current version, it contains 38 variables of which 18 are obtained from survey data, while the others are based on objectively quantifiable data¹. The variables are grouped into the following five major areas, for which individual sub-indices are calculated: size of government, legal system and property rights, sound money, freedom to exchange with foreigners, and regulation of credit, labour and business. After experimenting with different weighting approaches, in its recent version Gwartney and Lawson (2003b) calculate the aggregate index simply as the unweighted mean of its five sub-components. For each

¹ Survey data come from the International Country Risk Guide and the World Economic Forum's Global Competitiveness Report.

variable, economic freedom is measured on a scale of 0 to 10 where 0/10 means that a country is completely unfree/free.

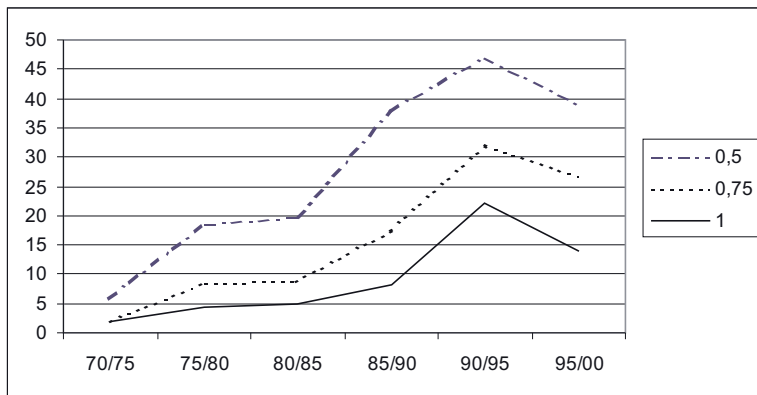
Although there are necessarily many debatable issues in the context of such an ambitious indicator, the EFW indicator offers a good starting point for a quantification of economic reform processes. Its availability for a large country sample since 1970 makes it preferable to the Heritage Foundation's index of economic freedom (O'Driscoll et al. 2003), which only covers the years since 1995. The availability of sub-indices is of particular interest in the context of economic reform obstacles. It makes it possible to test for the impact of potential obstacles on different fields of economic reform. A further big advantage is the fact that a wide literature has established at least a correlation if not a causal relationship between the indicator and economic growth (for a survey, see de Haan 2003). Measuring reform processes on this basis therefore implies the measurement of structural changes conducive to growth.

For the analysis, a reform event is defined as a "significant" change in the EFW index within a five-year interval where "significant" stands for an increase of 0.5, 0.75 or 1 index point. The focus on the binary variable (reform yes or no) corresponds to the nature of reform blockades in which a sweeping reform breakthrough is needed, associated with a wide and significant adjustment of outdated structures.

Figure 1 and Tables 1–2 describe some characteristics of the reform history since the seventies. Reform activities started to accelerate in the eighties and reached a peak in the 1990–1995 period. This peak period was also characterised by particularly ambitious projects with a large share of "big" reforms (increase in EFW indicator by at least one point) in total reform events. Over time, a marked shift in reform priorities has occurred: during the 70s and 80s, the reform focus was on freedom of exchange, legal system/property rights, and stable money. The size of government has become a major reform issue only since the second half of the eighties. Regulation was addressed even later, starting with the peak reform period 1990/95. Reform history differs among income groups: Low-income countries experienced a delay in following the global reform trend with rising numbers of reform events only since 1990/95. Reform frequency in high-income countries peaked relatively early in 1985/90.

As explained in our theoretical considerations in Section 2, at least three classes of variables are potentially helpful for explaining the occurrence of reforms: first, variables representing the objective need for reforms; second, variables that are proxies for relevant political-economic factors; and third, variables that could be related to phenomena of rational ignorance and irrationality. The specific selection of variables in the following is influenced by data availability, whereby extensive use has been made of the World Bank's World Development Indicator (WDI) data base (see Appendix for variable definitions and sources). The WDI data has the

Figure 1: Frequency of Reform Events (in % of Period Observations)



advantage of covering the time span and country sample of the Fraser Institute's economic freedom indicator and thus can be integrated into the analysis.

A reform event is defined as a change in the EFW aggregate indicator by at least 0.5 / 0.75 / 1.0 points.

Table 1

**Frequency of Reform Events (EFW Sub-Indicators):
in % of Period Observations (Total Number of Period Observations)**

		Period					
		70 / 75	75 / 80	80 / 85	85 / 90	90 / 95	95 / 00
							All periods
Change of at least 0.5 in sub-index	Government	15.6 (96)	13.1 (107)	23.4 (111)	46.5 (114)	42.1 (114)	33.6 (122)
	Legal system	10.0 (50)	70.0 (50)	27.8 (90)	27.3 (110)	47.7 (111)	22.0 (123)
	Money	11.2 (107)	30.4 (112)	46.4 (112)	27.6 (116)	38.0 (121)	55.3 (123)
	Exchange	39.7 (73)	21.3 (94)	22.5 (102)	28.6 (105)	50.0 (106)	29.9 (117)
	Regulation	15.6 (45)	9.6 (73)	9.8 (102)	10.4 (106)	41.5 (118)	26.8 (123)
							(567)

Table 2

**Frequency of Reform Events for Country Income Groups (EFW Indicator):
in % of Period Observations (Total Number of Period Observations)**

		Period						
		70/75	75/80	80/85	85/90	90/95	95/00	All periods
Change of at least 0.5 in EFW index	High-income countries	0.0 (26)	17.9 (28)	20.0 (30)	60.0 (30)	30.0 (30)	19.4 (31)	25.1 (175)
	Upper-middle-income countries	0.0 (9)	38.5 (13)	20.0 (20)	36.4 (22)	54.5 (22)	42.3 (26)	35.7 (112)
	Lower-middle-income countries	18.2 (11)	6.7 (15)	20.8 (24)	40.7 (27)	72.4 (29)	42.4 (33)	38.8 (139)
	Low-income countries	14.3 (7)	14.3 (14)	17.9 (28)	13.3 (30)	34.4 (32)	51.5 (33)	27.8 (144)

World Bank income classifications

Proxies for need of reforms:

Relevant variables to be included in this class are unemployment and growth – the latter defined as growth of total GDP or GDP per capita. In line with the hypothesis that a crisis is helpful to overcome reform blockades, changes in growth and unemployment rates should be more relevant than levels.

Variables describing a country's openness like export ratios and the extent of capital flows should also play a role, since they indicate the intensity of present external competitive pressure. The same holds for the change of the real effective exchange rate, which signals the development of external competitiveness. Sometimes the size of a country's population is also interpreted as a proxy for external competition in the sense that a country with a small population should suffer from losing competitiveness faster than a big country.

Moreover, a natural proxy for the economic need of reforms is the starting value of the economic freedom indicator: The closer the distance to the maximum value of 10, the lower the need for reform activities according to the logic of this study's approach.

Proxies for political-economic reform resistance

Variables in the WDI data set that may be used as political-economic proxies are mainly those related to the population's age structure. As argued in Section 2, an important element of rational reform resistance relates to the discounting of future reform benefits. This discounting is likely to increase with age. A variable such as population share above 65 should, therefore, be negatively linked to reform cap-

ability. Similarly, a negative link can be expected in regard to dependency ratios. A positive link should exist between reform capability and life expectation: the longer the individual time horizon, the more attractive the reforms associated with J-curve effects.

Other variables that would be useful, such as indicators of inequality, are not available for a reasonable number of observations.

Proxies for rational ignorance and/or limited rationality

Reform resistance associated with rational ignorance or innate irrationalities should be a negative function of the population's level of economic education. People who are knowledgeable about the economic system should also be more competent judges of appropriate reforms at the ballot box. Even though they have no direct incentive to inform themselves or to judge in a rational way, a high level of economic education should – as a by-product – lead to better-informed and more rational voting decisions. Unfortunately, there is no internationally consistent data set on economic education. Therefore, variables describing the general state of education such as school enrolment ratios have to be used. Also WDI data quantifying the availability of media information (newspapers, televisions) are experimented with – although it is not clear whether these very general media indicators are really a good indicator of a country's quality of media information. A more direct proxy for the population's economic education are variables quantifying the development of the financial sector such as credit aggregates – the presumption here is that a correlation exists between a country's financial development and the population's economic knowledge.

Table 3 presents the relevant variables' means separately for reform and no-reform country-period observations together with the results of the t-test for a difference in means. The variables are measured prior to the reform/no-reform period (for details of calculation, see Appendix) to avoid impacts of reversed causality (for example from reform activity to growth).

Prior to a reform period, countries have the following significant (at least 10% significance for t-test) differences compared to countries prior to a period without major institutional changes: lower economic freedom, lower growth, and smaller increase in p.c. income (PPS) in preceding years.

Many other differences – though not statistically significant – show expected signs. Reform countries, compared to no-reform countries, are characterised by a larger growth reduction, a smaller p.c. income (PPS), a higher unemployment rate, a larger increase in unemployment rate, a larger real appreciation of the exchange rate, smaller dependency ratios, a longer life expectation, a smaller share of population below 65 and higher school enrolment ratios (tertiary schooling).

Table 3
Test for Differences in Means for Reform / No Reform Events
(Number of Observations in Brackets)

	Reform (EFW increase of at least 0.5)		Significance
	Means prior to reform period	Means prior to no-reform period	t-test
proxies reform needs			
Start value EFW indicator	4.98 (178)	5.66 (392)	0.000
Growth rate	2.82 (176)	3.87 (376)	0.000
Change in growth rate	-0.92 (171)	-0.53 (366)	0.292
GDP p.c. (PPS)	5703.35 (171)	6359.91 (328)	0.227
Change in GDP p.c. (PPS)	1.26 (153)	1.35 (270)	0.001
Unemployment rate	8.73 (94)	7.65 (135)	0.173
Change in unemployment rate	1.39 (63)	0.64 (91)	0.209
Change in real effective exchange rate	0.99 (80)	0.95 (142)	0.448
Private capital flows in % GDP	11.37 (158)	16.02 (294)	0.130
Exports / GDP	31.95 (176)	34.12 (372)	0.319
Size of population	41.51 (177)	37.95 (387)	0.748
Proxies demographics related reform resistance			
Dependency ratio	0.71 (177)	0.74 (387)	0.150
Life expectation	65.24 (177)	63.85 (387)	0.131
Population share > = 65	6.32 (177)	6.60 (387)	0.466
Proxies rational ignorance / limited rationality			
Daily newspapers per 1,000 people	122.44 (170)	135.91 (377)	0.329
TV-sets per 1,000 people	171.89 (174)	183.33 (367)	0.464
School enrolment ratio (tertiary education)	16.00 (168)	15.13 (369)	0.501
Domestic bank credit in % GDP	52.35 (172)	54.86 (374)	0.488

Variables with a “wrong” and insignificant sign are: capital flows, export ratios, number of newspapers and televisions and bank credit to the private sector. Population size is not significantly different between reform and no-reform countries.

Overall, the proxies for the need of reforms (in particular, the growth variables) perform best in this simple testing. Important proxies for political-economic reform resistance (like age structure and dependency ratio) and for limited rationality/information (school enrolment) show the correct sign but lack significance. Obviously, this kind of simple bivariate testing only allows limited insights. Furthermore, due to the fact that the comparison is based on country-period-observations, one country can be included among the reform observations several times,

which opens the door for country-specific unobserved effects. Furthermore missing variables lead to different numbers of observations for each reform factor, which again could bias the results. Hence, this kind of descriptive exercise must now be substantiated by a more refined multivariate analysis.

4. Econometric Analysis

The next step of the analysis is to explain the occurrence of reform events in the framework of a probit model. The binary 1/0 variable reform/no-reform is now modelled to be jointly determined by a whole range of explanatory variables that – according to our theory – could have an impact on the probability of reform. In order to avoid difficulties with reversed causation, explanatory variables are measured prior to the five-year-window which is the basis for the observation of a reform event. Generally, the estimations include time and income group dummies in order to control for significant differences in reform behaviour across income groups and periods.

Compared to Section 3, data availability leads to exclusion of a number of variables from the regressions: unemployment ratio, GDP p.c., and real exchange rates. Furthermore, a number of variables turned out to be insignificant and have, therefore, been excluded from the regressions presented: dependency ratio, relative number of newspapers and TV-sets, size of population, bank credit to the private sector.

Nevertheless, proxies of each of the groups “reform needs”, “rational resistance”, and “limited information/rationality” survive and contribute to the explanation at least of certain aspects in the determination of reforms.

With regard to the overall reform activity (reform event defined as significant change of aggregate EFW indicator, Table 4) the following results emerge: a number of period and income group dummies are highly significant. Period dummies hint at a particularly reform-friendly mood since 1990/95. *Ceteris paribus*, high-income countries are significantly more reform-friendly than lower-middle and low-income countries. Economic growth rates and the start value of economic freedom have a significant impact and support the view that economic necessity is a relevant driving force for reforms. Economic openness measured by the export ratio has the correct sign in the sense that greater openness increases the pressure for reform but reaches 10 percent significance only in the regression for substantial reform events (change of EFW indicator of at least 1 point). Age structure as proxy for rational reform-resistance of population groups with heavy discounting turns out to be weakly significant. The same holds for life expectation, although only in one specification. In contrast, school enrolment as a proxy for limited rationality is insignificant in all three specifications. Thus, the regressions for overall reform indicators do not suggest relevance of limited rationality.

Table 4

Probit Regression: Reform Determinants, EFW Total Indicator

Dependent variable: reform event yes (1) / no (0) Periods: 70 / 75 – 95 / 00 Definition of reform event: increase in EFW indicator of at least:			
Independent variables	0.5	0.75	1
Constant	1.306 (0.24)	0.765 (0.55)	3.256*** (0.03)
Dummy 75 / 80	0.251 (0.49)	0.383 (0.49)	-0.078 (0.90)
Dummy 80 / 85	0.427 (0.22)	0.606 (0.26)	0.198 (0.73)
Dummy 95 / 90	1.116*** (0.00)	0.931* (0.08)	0.474 (0.40)
Dummy 90 / 95	1.553*** (0.00)	1.667*** (0.00)	1.429** (0.01)
Dummy 95 / 00	1.585*** (0.00)	1.794*** (0.00)	1.309** (0.02)
Dummy upper middle income	-0.572** (0.03)	-0.276 (0.34)	-0.539 (0.12)
Dummy lower middle income	-0.852*** (0.00)	-0.737** (0.03)	-1.167*** (0.00)
Dummy low income	-1.264*** (0.00)	-1.092** (0.03)	-2.052*** (0.00)
Start value EFW	-0.661*** (0.00)	-0.630*** (0.00)	-0.769*** (0.00)
Growth	-0.015 (0.49)	-0.065** (0.01)	-0.080*** (0.01)
Export-GDP-ratio	0.001 (0.73)	0.001 (0.72)	0.008* (0.06)
Life expectation	0.027* (0.08)	0.025 (0.15)	-0.002 (0.90)
Population share >65	-0.053* (0.05)	-0.053* (0.09)	-0.039 (0.28)
School enrolment	-0.001 (0.84)	-0.008 (0.37)	-0.006 (0.59)
Obs.	519	519	519
Reform obs.	166	98	57
R2 (Mc Fadden)	0.21	0.25	0.30

p values in parentheses, * / ** / ***: significant at 10% / 5% / 1%.

This picture changes considerably when different reform fields are analysed separately on the basis of the five EFW sub-indicators (Table 5). First of all, income group dummies lose significance – income levels do not seem to affect reform probabilities when the approach differentiates between reform fields. Time dummies, however, remain significant and repeat the message of Table 1 that the focus of reforms shifts over time. There appear to be changing “fashions” of reform independently from the objective economic environment – for example the fact that cutting back government became a trend since 1985 / 90 or that the period 1990 / 1995 was the heyday of deregulation.

Proxies for reform needs lose significance in these differentiated regressions. While the start value of the EFW indicator continues to be highly significant, the growth rate keeps its expected sign but is now insignificant for each sub-indicator. Export openness is significant with the correct sign in the context of reforms targeted at freedom of exchange, which is a plausible result. The larger the shares of

Table 5

Probit Regression: Reform Determinants, EFW Sub-Indicators

Dependent variable: reform event yes (1) / no (0) Periods: 70 / 75 – 95 / 00 Definition of reform event: Increase in EFW sub-indicator of at least 0.5:					
Independent variables	Government	Legal structure	Exchange	Money	Regulation
Constant	-0.449 (0.67)	-0.940 (0.41)	1.851* (0.08)	1.472 (0.17)	0.227 (0.86)
Dummy 75 / 80	-0.323 (0.38)	1.692*** (0.00)	0.373 (0.21)	-0.581** (0.04)	-0.060 (0.88)
Dummy 80 / 85	0.373 (0.24)	0.656** (0.04)	0.860*** (0.00)	-0.714*** (0.01)	-0.139 (0.71)
Dummy 85 / 90	0.940*** (0.00)	0.598* (0.07)	0.503* (0.08)	-0.546** (0.05)	-0.304 (0.44)
Dummy 90 / 95	0.830*** (0.01)	1.277*** (0.00)	0.774*** (0.01)	0.309 (0.26)	0.982*** (0.01)
Dummy 95 / 00	0.668** (0.05)	0.589* (0.09)	1.376*** (0.00)	-0.129 (0.66)	0.634* (0.09)
Dummy upper-middle-income	0.057 (0.83)	-0.538** (0.05)	-0.032 (0.90)	-0.139 (0.61)	-0.106 (0.71)
Dummy lower-middle-income	0.127 (0.67)	-0.322 (0.29)	-0.123 (0.66)	-0.302 (0.32)	-0.050 (0.88)
Dummy low-income	0.342 (0.43)	-0.397 (0.39)	-0.512 (0.23)	-0.184 (0.68)	-0.523 (0.31)
Start value EFW	-0.267*** (0.00)	-0.245*** (0.00)	-0.296*** (0.00)	-0.324*** (0.00)	-0.302*** (0.00)
Growth	-0.019 (0.37)	0.008 (0.72)	-0.010 (0.64)	-0.007 (0.74)	-0.014 (0.59)
Export-GDP ratio	0.004 (0.17)	0.003 (0.33)	0.006** (0.05)	-0.009** (0.02)	0.005 (0.17)
Life expectation	0.009 (0.55)	0.017 (0.27)	-0.019 (0.18)	0.013 (0.35)	-0.003 (0.87)
Population share > 65	-0.038 (0.15)	-0.015 (0.59)	-0.006 (0.83)	-0.053* (0.07)	0.040 (0.19)
School enrolment	0.018*** (0.00)	0.005 (0.48)	-0.001 (0.91)	0.000 (0.95)	0.010 (0.16)
Obs.	515	481	519	500	491
Reform obs.	159	155	207	163	99
R2 (Mc Fadden)	0.12	0.13	0.10	0.13	0.16

p values in parentheses, * / ** / ***: significant at 10% / 5% / 1%.

exports in GDP, the more costly restrictions to international transactions are. The export ratio's significance in regard to stable money with a negative sign is less plausible, however.

Similar to the reform need proxies, the variables representing rational reform resistance perform worse in these differentiated regressions: life expectation and age structure are always insignificant with one exception in the stable money regression.

The contrary holds for the proxy for the extent of rational ignorance and limited irrationality. School enrolment is now strongly significant with the expected sign in the reform regression for the size of government. Countries with better-educated citizens are more likely to reform the government sector and to cut back government spending and taxes.

5. Conclusions

There is no one-dimensional answer why countries have different inclinations to delay or even refuse growth-inducing reforms. Traditional answers given in the economic literature are related to the extent of economic crisis or to political-economic factors. Our theoretical considerations suggest that these traditional answers are possibly incomplete. Limited rationality on the side of voters might be a further factor in the explanation of reform deadlocks: blockades might occur because it is individually costless not to judge reforms in a completely rational way.

Our empirical approach is not guided by the ambition to explore the full universe of rational and irrational determinants of reforms. However, we try to shed light on the weight of three classes of factors: those related to the objective need for reforms, those representing political-economic reform-resisting forces, and those associated with rational ignorance or limited rationality.

Not surprisingly, there is strong evidence that the objective need for reforms is highly relevant. The worse a country's initial institutional setting, the more likely reforms are. Equally, poor growth performance has a positive impact on a country's ability to overcome reform resistance.

Among political-economic proxies, the age structure has a certain explanatory power, at least in the regressions based on the EFW total indicator. Here, reform probabilities decrease with the age of a population. This is a worrying result given the demographic future of industrial countries and the resulting needs for reforming the social systems.

Data availability does not make it possible to test for the relevance of limited information and limited rationality in an extensive way. School enrolment as the related proxy is significant only in the government reform regression. This means a higher level of education makes it easier to reform the government sector. This is

a plausible result: for example, cutting back government subsidies is associated with immediate and highly visible losses, whereas the benefits (future tax cuts, growth) are more abstract. A higher level of education may make it easier to welcome this kind of institutional change.

Apart from these results, other interesting insights emerge as a by-product: The significant period dummies are evidence of changing fashions in the focus of reforms. Deregulation, for example, only became a trend in the first half of the nine-ties.

Obviously, this empirical approach is just a first step and much more needs to be done to really understand the economics of reforms. Apart from macro-approaches as applied in this paper, micro-approaches based on experimental research seem to be necessary. Differencing, for example, between by poor information and limited rationality as causes for the rejection of beneficial reforms is hardly possible on the basis of macro data. In this sense, a frequent final conclusion applies particularly well to this paper: more research is needed on the empirics of reform.

References

- Alesina, A./Drazen, A.* (1991): "Why Are Stabilizations Delayed?", *American Economic Review*, 81 (5), 1170–1188.
- Caplan, B.* (2001a): "Rational Irrationality and the Microfoundations of Political Failure", *Public Choice*, 107 (3/4), 311–331.
- (2001b): "Rational Ignorance versus Rational Irrationality", *Kyklos*, 54 (1), 3–26.
- Daniel, K./Hirshleifer, D./Siew Hong Teoh* (2002): "Investor Psychology in Capital Markets: Evidence and Policy Implications", *Journal of Monetary Economics*, 49 (1), 139–209.
- De Haan, J.* (2003): "Economic Freedom: Editor's Introduction", *European Journal of Political Economy*, 19, 395–403.
- Downs, A.* (1957): "An Economic Theory of Democracy", New York: Harper.
- Fernandez, Raquel/Rodrik, D.* (1991): "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty", *American Economic Review*, 81 (5), 1146–1155.
- Gwartney, J./Lawson, R.* (2003a): "The Concept and Measurement of Economic Freedom", *European Journal of Political Economy*, 19, 405–430.
- (2003b): "Economic Freedom of the World", 2003 Annual Report, Fraser Institute, Vancouver.
- Heinemann, F.* (2001): "Die Psychologie begrenzt rationaler Wirtschaftspolitik: Das Beispiel des Reformstaus", *Zeitschrift für Wirtschaftspolitik*, 50 (1), 96–110.
- Kahneman, D./Knetsch, J. L./Thaler, R. H.* (1991): "Anomalies, the Endowment Effect, Loss Aversion and Status Quo Bias", *Journal of Economic Perspectives*, 5 (1), 193–206.

- O'Driscoll, Jr., G. P. / Feulner, E. J. / O'Grady, M.A.* (2003): "2003 Index of Economic Freedom", Heritage Foundation, Washington, DC.
- Pitlik, H. / Wirth, S.* (2003): "Do Crises Promote the Extent of Economic Liberalization? An Empirical Test", *European Journal of Political Economy*, 19, 565–581.
- Rodrik, D.* (1996): Understanding Economic Policy Reform, *Journal of Economic Literature*, 34 (March), 9–41.
- Samuelson, W. / Zechkauser, R.* (1988): "Status Quo Bias in Decision Making", *Journal of Risk and Uncertainty*, 1, 7–59.
- Tversky, A. / Kahneman, D.* (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model", *The Quarterly Journal of Economics*, 106 (4), 1039–1061.
- Wittman, D.* (1995): "The Myth of Democratic Failure: Why Political Institutions are Efficient". Chicago: University of Chicago Press.

Appendix: List of Variables

Variable	Calculation	Source
Growth	Average annual real GDP growth in five-year-period prior to observation period	WDI
Change growth	Difference average growth between two five-year-periods prior to observation period	WDI
GDP p.c. (PPS)	GDP p.c. in PPS (current international \$), in the first year of observation period	WDI
Change in GDP p.c. (PPS)	Change in GDP p.c. (PPS) during five-year-period prior to observation period	WDI
Unemployment rate	Unemployment rate at first year of observation period	WDI
Change unemployment rate	Five-year change in unemployment rate prior to observation period	WDI
Export-GDP-ratio	Exports in % GDP, in the first year of observation period	WDI
Capital flows	Private capital flows (in- and outwards) in % GDP, first year of observation period	WDI
Change real effective exchange rate	Change real effective exchange rate within five-year-period prior to observation period (decline stands for depreciation)	WDI
Population	Total population at first year of observation period	WDI
Population share > = 65	Population 65 and above in % of total population at first year of observation period	WDI
Dependency ratio	Age dependency ratio (dependents to working age population) at first year of observation period	WDI

Variable	Calculation	Source
Life expectation	Life expectation at birth at first year of observation period	WDI
Credit	Domestic credit provided by the banking sector, in % GDP, first year of observation period	WDI
School enrolment	School enrolment, tertiary (% of age group), in the first year of observation period	WDI
Daily newspapers per 1,000 people	First year of observation period	WDI
TV sets per 1,000 people	First year of observation period	WDI

“Observation period” refers to five-year-window for which the change in the EFW index is measured and, consequently, a country-period-observation is classified as reform or no-reform observation; WDI: World Bank World Development Indicators.

Explaining Reform Deadlocks

Comment

By Mathias Erlei*

Heinemann's innovative and interesting paper aims to analyze obstacles to efficiency enhancing economic reforms. This is both an interesting and important question of economic research. According to the theoretical literature the author distinguishes three groups of variables that plausibly have an influence on the probability of efficient reform programs: The first group tries to cover the need for economic reforms, the second covers distributional conflicts and the third group tries to cover aspects of limited rationality.

After having carried out a descriptive analysis of the data the author continues with an econometric analysis of his data set. In his aggregate estimation time dummies, income group dummies, the start value of the EFW index ("Economic Freedom in the World index") are highly significant factors in most of the estimations. Growth of GDP, the ratio of exports to GDP, people's life expectations and the population's share of people older than 65 years are only significant in a subset of the estimations. In his estimations with regard to the sub-indices, time dummies and the start value of EFW are again highly significant. The export to GDP ratio, the population's share of people older than 65 and school enrolment are only significant in some of the estimations.

Finally, the author draws some conclusions: (1) limited rationality seems to be particularly important in reforming the government sector; (2) there seem to exist "fashions" in the focus of reforms (time interval dummies) and (3) in older populations, reform probabilities tend to decrease.

I think this paper has four major strengths. First, the author makes a step in the direction of a particularly important research question that has, by and large, been ignored by empirical research. Second, I fully agree with the author that the EFW index is a very useful measure of the diverse aspects of the economic order. Third, it is extremely useful to take the change of EFW as a proxy variable of economic reform. I strongly recommend proceeding on this path of research. Finally, the author gets some interesting and intuitively plausible results.

* Institut für Wirtschaftswissenschaft, TU Clausthal, Julius-Albert-Str. 2, 38678 Clausthal-Zellerfeld, E-mail: m.erlei@tu-clausthal.de

Unfortunately, there are also some severe shortcomings of Heinemann's analysis. Firstly, Heinemann's conclusions seem to go slightly beyond his empirical results. When summarizing that "a higher level of education makes it easier to reform the government sector" he should add explicitly that his data also seem to show that it is only the reform of the government sector in which education seems to be relevant! This, however, is an implausible result: Why should such limited rationality only be relevant for government sector reform? There is no sensible theory available to me that makes such a result convincing. Remember that limited rationality, as the author explains it in the theoretical section, covers such broad aspects as "rational ignorance" (i.e. it does not pay off to be well informed) and other forms of bounded rationalities like status quo biases, endowment effects and loss aversion. All these arguments do not at all seem to be restricted to government sector reforms.

The author's second conclusion that there exist *fashions* in the focus of reforms is also rather questionable. One can as well argue that the remarkable significance of time interval dummies might be explained by different structural needs for economic reform in different periods according to global changes in economic problems which are not covered by the author's variables. Finally, life expectation is only significant in one out of eight estimations and only at the 10 percent level. I doubt whether this is sufficient to submit a relationship between reform probability and population's age.

I think that there is another very important problem with the adequate treatment of limited rationality in the author's regressions. Let us first look at "rational ignorance". Obviously, it does not pay off for an individual to get information about the reform proposals. However, this problem exists in all countries and every time. Therefore, it should be *covered by the constant* in the regression estimations. Furthermore, the relation between education and "rational ignorance" is loose at best! However, if one really tries to measure "rational ignorance" one should take variables such as the percentage of people knowing the most important politicians and/or policy proposals.

Next, we should discuss whether limited rationality is treated in an adequate way. In Section 2 the author explains that bounded rationalities refer to the important findings of behavioral economics. In particular, he lists arguments "such as the 'status quo bias', the 'endowment effect' and 'loss aversion'". Again, are these aspects of limited rationality restricted to countries and time periods? If not, then they are again *covered by the constant* of regression estimations! If so, they should be covered by adequate proxies. Unfortunately, "education" is almost certainly not a good proxy for such effects as the author has listed them in the paper: as far as I know, status quo biases, endowment effects, and loss aversion effects also exist for economic graduates. For example, sensible proxies for "loss aversion" might be the percentage of entrepreneurs in the population or the magnitude of the stock market in comparison to the securities market. Obviously, the generation of these

data may be very time consuming and costly. However, it seems to be worthwhile because the underlying research question is extremely important for actual economic policy.

Without doubt, population age is most important for the problem of pension systems. Unfortunately, however, it presumably is of only minor importance to other kinds of reform. As the author tries to cover the whole range of economic reform, his choice of distributive variables, i.e. people's life expectation and the population percentage of those people who are older than 65, is highly incomplete. What is really needed is an indicator of society's degree of organized interests. Such interest groups are active mostly with regard to their own sectors. Furthermore, they usually have an information advantage which makes it possible for them to block several kinds of reform. As each interest group concentrates on its own sector, each interest group dominates its own area. This gives additional power to the interest groups. It is obvious that Heinemann's variables do not cover such distributional aspects at all.

Finally, two arguments from "standard theory", risk aversion under uncertainty and the "war of attrition" mechanism, are not covered at all. In general, I have to criticize that there is only a *very weak* link between theory that is discussed in the paper and the empirical part. Consequently, the empirical sections of the paper do at best give some vague hints about the relevance of arguments discussed in the theoretical section.

Summarizing, although I am rather critical of several details in Heinemann's empirical analysis, I wish to express my accord with his principle approach to measure reform activities by changes in the EFW index. As far as I know, this is innovative and it certainly is the most convincing way to approach the problem that I know.

The Reform of Local Public Services of General Interest in Europe

By Peter Haug and Martin T. W. Rosenfeld*

Abstract

The benefits of a reduced supply of local public services may more than outweigh the supposed welfare losses. This has been suggested by various theoretical and empirical investigations in many fields of economics during recent decades. Nevertheless, local and national politicians, trade unionists, charities, and other lobbyists have succeeded in preventing further liberalisation of services of general interest in Europe. This paper examines why these agents of preservation have been and are still successful. The analysis is based on an institutional economic approach. Several policy measures and institutional changes are suggested to either reduce influence of agents of preservation or to compensate them for their losses.

JEL Classification: D78, L33, H42

Keywords: economics of reform, local public goods, services of general interest, liberalisation, interest groups, efficiency

1. Introduction

In the 1980s and 1990s, there was a strong tendency towards “privatisation” of local public goods and services in many industrialised countries. Local authorities wanted to increase administrative efficiency, and the higher levels of government, especially the EU level, were exerting political pressure to liberalise economic structures and harmonise these structures within the EU. While the liberalisation process has produced some significant results¹, in several countries in recent years, for example Germany, shifting of local economic activities to the private sector has not made any further progress. At the German federal level, there have been, up to now, only a few initiatives and discussions on further reforms of local public

* *Contact address:* Halle Institute for Economic Research, Department for Regional and Urban Economics, Kleine Märkerstr. 8, D-06108 Halle (Saale), Germany. Phone +49(0345)7753-750 or -709, Fax: +49(0345)7753-876, Email: Peter.Haug@iwh-halle.de, Martin.Rosenfeld@iwh-halle.de.

¹ This is apart from the liberalisation of the electricity and the gas markets, which have been monopolies under local public or regional public control.

services. And at the EU level, the European Parliament has been struggling (at least since 2001) against the initiatives of the European Commission to restrict economic activities of the local sector according to the rules of competition for the single market. In January 2004, parliament passed a resolution in which the parliamentarians strongly opposed the commission's ideas expressed in the Greenbook "Services of General Interest" (May 2003).

In many cases, the "privatisation" of local public services has been only a formal one. Local authorities have just "outsourced" the production of local public goods and services to special-purpose firms that operate according to the rules of private law despite being controlled more or less totally by the local government. Local authorities and federations thereof justify this behaviour and the protection of the current level of production against all efforts towards further liberalisation by claiming that the goods and services in question are local *services of general interest*, and that securing the supply of these goods and services must be an "eternal" task of the local authorities. Other pressure groups support this theory. So far, all these "*agents of preservation*" have been more successful than those who are in favour of rolling back the local public influence on the economy ("*agents of change*").

The paper will first discuss the expected benefits and the supposed costs of a comprehensive privatisation in the field of local public services of general interest (see Section 2). Today, public production units are responsible for a wide range of rather different services, from water supply and waste disposal to public libraries, street cleaning, and childcare services². Against the background of the discussion in Section 2, Section 3 will answer the question of why the expected increase in welfare has not yet taken place, and why the agents of preservation have been able to stop institutional change up to now. This section of the paper will concentrate on the conditions in Germany. In the last section, we will ask what could be done in order to overcome the resistance of the agents of preservation.

2. The Benefits of a Reduced Supply of Local Public Services

In this section we discuss how efficiency and common welfare, as well as equity, are affected by economic activities in the local public sector. The analysis will follow the argumentation used in the relevant economic literature on regulation. Deviating from prior work, we apply these considerations to local goods and services. In addition, we discuss arguments put forward frequently in favour of maintaining local governments' current economic activities.

² The discussion will not refer to services generally produced at the state or federal level, nor to services that in Germany are called "services for securing the sovereignty of the state", for example those of registry offices. In addition, we will ignore the problem of public production of services for which the localities have no monopoly, for example car washing.

2.1 General Aspects of Welfare Economics of Local Public Activities

An accepted truth of traditional welfare economics is that aside from certain cases of market failure, private enterprises are most efficient at producing goods and services. It should be pointed out that it is useful to distinguish between efficient *provision* and efficient *production*. The first means providing a certain output of goods and services according to the preferences of households or enterprises, while the latter involves producing a given output at minimum costs. It should be noted further that, in this context, private production and private provision stand for provision or production by enterprises owned solely by private shareholders. Private provision and production do not refer to production and provision of local public enterprises of legal form under private law³.

The most relevant cases of market failure at the local level are externalities, public goods, natural monopolies, and asymmetric information. Provision of sewage treatment, waste disposal services, drinking water, and hospitals involves *positive externalities*. Even households without access to these services benefit from a reduced risk of infectious diseases and reduced annoyances of uncontrolled waste and sewage disposal. Fire brigades protect buildings in the neighbourhood of a fire as well. It is hardly possible⁴, and not desirable from a welfare point of view, to exclude anyone from the positive spillovers. Thus, insufficient internalisation of positive externalities will cause an undersupply of the specific goods and services.

Another source of market failure are the *natural monopolies* common in network-related industries. At a local level, especially pipes for distributing water and sewage are characterised by economies of scale and decreasing average costs⁵. Therefore, doubling or tripling networks would be a highly inefficient waste of economic resources. Furthermore, the costs of installing pipeline networks are mostly “sunk costs” because mains, pumping stations, service connections, and sewers can hardly be used for alternative purposes. High sunk costs in these industries help to erect further barriers to market entry which only permit competition *for* the market, but prevent competition *within* the market. Although many of the goods and services mentioned are considered to be “traditional” tasks of local governments, at least by certain lobbyists, existing market failures may only de-

³ The most common legal form (under private law) of local public enterprise in Germany is the “Gesellschaft mit beschränkter Haftung” (GmbH, corporation with limited liability of shareholders), which is (roughly) similar to the British “limited company” or the French “SARL”. In order to simplify matters we will not deal with mixed enterprises, that is enterprises with varying percentages of public *and* private shareholders.

⁴ The costs of exclusion are prohibitively high.

⁵ Those empirical studies of US-water utilities by Kim (1985), in which 60 utilities were analysed by cross-section or Armstrong and Leppel (1996) (panel data analysis for nine investor-owned utilities) which found no empirical evidence for economies of scale, used aggregated total costs instead of estimating separate cost functions of distribution and production.

mand regulation of the relevant markets or public provision, but do not legitimate the current prevailing public production in most cases.

2.2 Arguments against Public Production of Local Public Goods and Services

Arguments for liberalisation of the local public sector are backed by some inherent internal inefficiencies in public enterprises which result from different control costs and incentive schemes of public enterprises, as compared with the private sector.

2.2.1 Principal-Agent Problems are More Relevant to Public Production

One of the most notorious efficiency killers in all categories of enterprises is imperfect and asymmetric information. A main characteristic of large private corporations is that managers are usually not owners of the company⁶. As utility-maximising “agents”, they have incentives to abuse their specific internal know-how for rent-seeking at the expense of their “principals” (the shareholders) by reducing effort or avoiding risky but highly profitable investments in order to make their jobs safer. For the principal(s), detecting shirking involves costs of control. There are some theoretical arguments that suggest control costs in public enterprises tend to be higher than in private enterprises and may cause higher production costs in the public sector⁷.

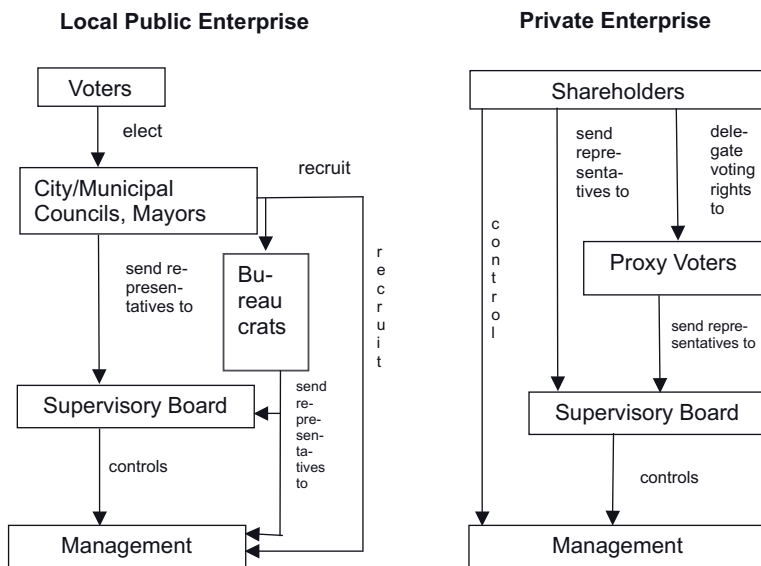
First of all, the control mechanisms in public enterprises are more complex and involve a higher degree of interconnections and multilateral dependencies between different levels. Owners of local public enterprises (private households living within the boundaries of a locality) have to delegate their rights to control by elections to the municipal councils and mayors. These institutions recruit employees (“bureaucrats”) for local public enterprises, as well as for tasks in public administration. In order to control top management, which often consists of bureaucrats, an obligatory controlling body has to be installed⁸. Supervisory boards are mainly formed by members of the municipal or city councils, mayors, unionists, and bureaucrats. Consequently, independent and competent economic evaluation of the

⁶ To put it precisely, shares or stock options are part of the remuneration of top managers, but their main source of income is a salary, which tends to vary only slightly with enterprise profits.

⁷ We partly follow Mühlenkamp’s argumentation. Refer to Mühlenkamp (1999), pp. 99–112 for more information.

⁸ This can be either a special committee of the municipal or city council (for example “Werksausschuss” in Germany) for local public enterprises under public law, or a supervisory board in corporations under private law.

management's business operations is not guaranteed. The following figure depicts the different control mechanisms of a purely (local) public and a purely private company:



Source: Following Mühlenkamp (1999, p. 100).

Figure 1: Stylised Multilevel Principal-Agent Relationships in Private and Public Enterprises

Voters are unable to exercise direct control of public enterprises and thus have to cope with high control costs combined with negligible changes of their personal utility because of their marginal share in public enterprises. Other citizens cannot be excluded from the benefits of improved control, which therefore can be characterised as a public good. The marginal share per voter in local public enterprises and non-excludability from the benefits of control favour insufficient controlling activities. On the other hand, even in big private corporations, the majority of the shares are often held by a small number of institutions such as financial institutions and other companies. Small shareholders (“smallholders”) can delegate their control rights to proxy voters, which enforces further concentration of management control. Furthermore, the average number of shareholders of small and medium-sized private enterprises is rather small. Thus, it can be concluded that owners of private enterprises have better cost-benefit-ratios of management control than “owners” of public enterprises.

2.2.2 Incentive Problems for Public Managers

In addition to the arguments discussed above, some other reasons for decreased efficiency of local public enterprises have to be considered.

Missing Disciplining Effects of Capital Markets:

On perfect capital markets, share prices provide information about the discounted value of expected earnings of firms and therefore, to some extent, about the quality of management performance. Whether capital markets are perfect or not, this mechanism is not relevant for public managers, because voters can't sell their "shares" in public enterprises except by moving away from the community ("voting with their feet"). In other words, their transaction costs are prohibitively high. Takeovers in cases of bad management do not seem to be a very likely threat for managers of public enterprises. Public managers are often employed and dismissed according to political considerations⁹. Another relevant point is that public enterprises cannot go bankrupt. This fact grants them access to special loans ("Kommunalkredite") at more favourable conditions than private enterprises.

Missing Disciplining Effect of Competition:

The disciplining effect of competitive markets, which makes it possible to compare managers' performance, is much less relevant for public enterprises. Local public utilities in particular often act within protected monopoly markets. Therefore, evaluating performance indicators of local public enterprises in some kind of benchmark process can only be a second-best alternative. Benchmarking, in general, tends to compare "sloppiness with sloppiness".

Self-interest of public managers:

Niskanen (1971, 1978) showed in several models that bureaucrats whose utility depends on budget expansion or output quantities tend to produce either at higher costs or provide excessively large output quantities. Niskanen's theory can be applied to public enterprises as well, because their revenues, and therefore their budgets, are largely determined by political interventions. The majority of public enterprises depend on transfer payments to cover their deficits and are allowed to fix prices only with the permission of the responsible authorities¹⁰.

⁹ It is questionable whether hostile takeovers are realistic threats even for managers of big private corporations in Germany.

¹⁰ For more information, see Mühlenkamp (1999), p. 110.

2.2.3 Comparing Governance Costs of Public and Private Production in Natural Monopolies

Up to this point we have discussed arguments in this section which may recommend private production of local public goods and services even in cases of market failure. We will now turn to a special case which might be most relevant for local public utilities. Integrating governance costs might legitimate public production in the case of natural monopolies¹¹.

For illustrative purposes, a simple microeconomic model will be introduced. We will consider a single-product enterprise with sub-additive cost functions and decreasing average costs. The representative private and public producer face high “sunk costs” and produce and supply goods for a non-contestable local market. For simplicity, the cost functions are assumed to be linear with fixed costs a (short- or medium-term perspective) and constant variable (and marginal) costs b per unit:

$$(1) \quad K_{pub/pri} = a_{pub/pri} + b_{pub/pri} \cdot x \text{ with } a_{pub} > a_{pri} \text{ and } b_{pub} > b_{pri}$$

Fixed, marginal and average costs of production are assumed to be lower for private enterprises. Higher costs per unit result from insufficient incentives for managers and other employees or from distributive goals of public enterprises. It is further assumed that the rising costs, which can be assigned to efficiency losses caused by shirking and inefficient use of resources, more than outweigh lower capital costs of public enterprises.

Demand functions $D(x)$ for the specific good representing the “marginal willingness to pay” of the representative household for the x th unit are supposed to be independent of the supplier.

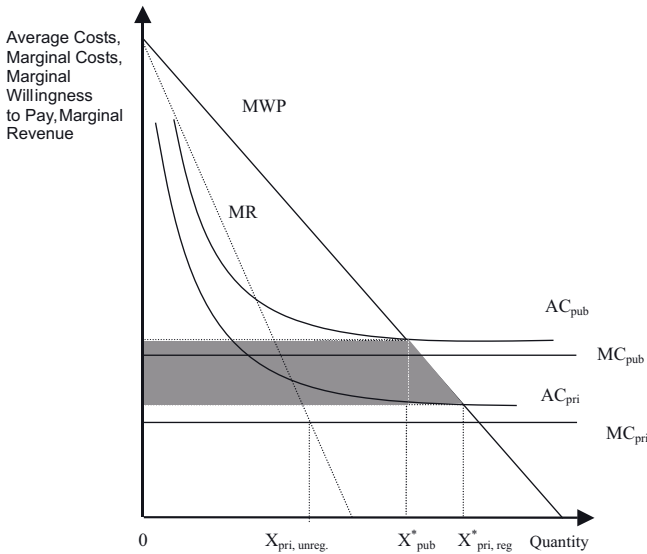
It is further assumed that not only private but also public providers are unwilling to supply welfare maximising quantities (here: a second-best solution with uniform pricing and total costs covered by revenues) in an unregulated market¹². Thus, regulating both private and public enterprises to make them supply the efficient quantity $x_{pri, reg}^*$ (instead of providing the Cournot-solution $x_{priv, unreg}^*$) and x_{pub}^* involves governance costs G_{pri} and G_{pub} . We assume G_{pri} to be higher than G_{pub} ¹³. Inserting the regulation costs into (1) yields the modified cost functions:

$$(2) \quad K_{pub/pri} = a_{pub/pri} + b_{pub/pri} \cdot x + G_{pub/pri}$$

¹¹ See Williamson 1985 or Mühlenkamp 1999.

¹² Niskanen's model of bureaucracy and the entire public choice literature – see for example Mueller (2003) – suggest that it is not very realistic to suppose politicians and their agents act like “benevolent dictators”.

¹³ G_{pri} represents the transaction costs or the costs of market use. G_{pub} is usually neglected in similar theoretical and empirical analyses of public “make-or-buy” decisions.



Source: Author's graphic.

Figure 2: Efficient Allocation of Private and Public Production in a Non-contestable Monopoly Market

In order to decide whether a specific good should be produced by a local public enterprise or bought from a private producer, it has to be analysed to determine which kind of production mode maximises social surplus (SSP). SSP, in a static sense, can be defined as total consumer surplus (consumers' total willingness to pay minus the actual price paid) and producer profits (actual price paid minus production and transaction costs) subject to equity constraints.¹⁴ Public production would be more efficient if its social surplus exceeded the social surplus of private production:

$$(3) \quad SSP_{pub} = \int_0^{x_{pub}^*} D(x)dx - K_{pub}(x) - G_{pub} > \int_0^{x_{pri,reg}^*} D(x)dx - K_{pri}(x) - G_{pri} = SSP_{pri,reg}$$

$$(4) \quad SSP_{pub} - SSP_{pri,reg} = - \int_{x_{pub}^*}^{x_{pri,reg}^*} D(x)dx - (K_{pub}(x) - K_{pri}(x)) - (G_{pub} - G_{pri}) > 0$$

$$(5) \quad SSP_{pub} - SSP_{pri,reg} = \int_{x_{pub}^*}^{x_{pri,reg}^*} D(x)dx - (a_{pub} - a_{priv}) - (b_{pub} \cdot x_{pub}^* - b_{priv} \cdot x_{pri,reg}^*) - \Delta G > 0$$

¹⁴ See for example Crew and Kleindorfer (1986), pp. 9–10.

The first and second term of (5) are negative. The sign of the third term is indefinite because $b_{pub} > b_{priv}$ and $x_{pub}^* < x_{pri, reg}^*$. Therefore, public production in the single-product case might be preferable to a regulated private monopoly only if the lower control costs in the local public enterprises more than outweigh the higher fixed and probably higher total variable costs plus the loss in aggregated willingness to pay. Figure 2 illustrates that the difference in costs of regulation has to exceed the grey area which represents the cost reduction for producing x_{pub}^* by a private producer and the net social surplus yielded by producing the x_{pub}^* th unit to the $x_{pri, reg}^*$ th unit by a private supplier.

As this simple static analysis of a single-product enterprise is based on some quite unrealistic assumptions, some further aspects ought to be integrated:

- *Quality matters:* Private and public suppliers choose some profit- or utility-maximising quantity-quality combination in order to evade price regulation. The potential results (increasing or decreasing quantities / quality standards after regulation) depend on information asymmetries and the relationship between quantity and quality (complements or substitutes) for the customers. This raises governance costs of private sector, as well of public sector firms¹⁵.
- *Multi-product enterprises:* Efficiency gains (or losses) from (dis)economies of scope by offering different goods (in Germany usually water, sewage treatment, electricity, gas, heating, and public transportation) by a single local public enterprise might lower (or raise) production costs. But potential private suppliers in the utility sector are mainly multi-branch enterprises and probably will also benefit from or be negatively affected by (dis-)economies of scope on a larger scale¹⁶.
- *Mixed enterprises:* Integration of private shareholders into public firms may increase efficiency (access to advanced technology, know-how transfer), but also raise costs of regulation.
- *Market failure versus government failure:* It makes sense to tolerate welfare losses in unregulated markets if gross welfare gains from regulation are low compared to the regulation costs or efficiency losses of public production. Net welfare losses of regulation may result from the fact that regulatory agencies are agents of elected politicians, which involves more similar principal-agent problems than for (local) public enterprises¹⁷.

¹⁵ For some examples of the theoretical literature dealing with quality aspects of regulation, see Sheshinsky (1976), Laffont and Tirole (1991) or Beitia (1995).

¹⁶ The German RWE corporation, for example, is a typical “multi-utility” enterprise. RWE’s core business includes electricity, gas, water, sewage treatment, waste disposal, and recycling services. For the theory of multi-product enterprises and the sustainability of natural monopolies, see for example Baumol (1977) or Bongaerts (1982).

¹⁷ Noll (1989) discusses several kinds of “government failure” in regulatory processes, resulting from interest group pressure (dividing rents) or information asymmetries between government and regulators.

2.3 Arguments Put Forward by Proponents of Retaining the Status Quo of Local Public Production

The factors discussed above are only to some extent suited to support the supposed superiority of (local) public production over private production. The huge amount of empirical privatisation, regulation, and outsourcing literature, however, provides some evidence for greater productivity of private enterprises or increases in productivity after privatisation¹⁸. Additional social, political, or environmental goals of public enterprises that mainly tend to increase labour costs cannot serve as a general excuse for internal inefficiency or wasting of tax revenues. The finding in some studies that competition (not ownership) is one of the main determinants of efficiency does not legitimate public production per se: for competitive markets it is difficult to justify any direct government intervention if no other severe market failure exists.

In the recent policy debate, some aspects of how local government enterprises might influence spatial factor allocation have been put forward as arguments for using public enterprises as tools of regional development policy in structurally lagging regions.

First of all, local public utilities often advance financial resources for other investments of local governments toward infrastructure. It is common practise for bigger local public corporations in Germany to pre-finance on most favourable terms, for example, modernization of old water conduits for other localities. One the one hand, German local governments are struggling for financial resources for necessary infrastructure investments, and without sufficient local basic infrastructure such as streets or utilities, attracting private firms and creating jobs is difficult. On the other hand, it is also true that both transfer payments and loans at interest rates below market rates might encourage overinvestment, for example when plants and equipment are replaced before the end of their useful lives or cheaper alternatives to replacement are neglected¹⁹.

Profits of some local public firms, especially in Germany, are considered to be a source of financial revenues for financing other local public goods and services. It is quite common to cross-subsidise public transportation by transferring profits from the local public utility sector. This practise has to be considered problematic from an allocation and distributive point of view. Charging profits of the utility

¹⁸ For an overview of recent international empirical studies on privatisation see Megginson and Netter (2001). Neal et al. (1996) compare productivity and profitability of investor-owned and government water utilities in California. A recent benchmark study of local water suppliers in Bavaria, see Rödl & Partner (2003), avoids analysing the impact of private involvement on internal efficiency of water utilities.

¹⁹ For example, the costs of relining old cast iron water mains with cement mortar are much lower than those of replacing them with new ones. However, local governments often seem to prefer the latter, as investment decisions of local public enterprises (especially German) are still dominated mainly by technical rather than economic considerations.

division against high deficits of the public transportation sector allows local public multi-product enterprises to minimise taxable profits. As no private enterprise can afford to keep subsidiaries with permanent deficits in the long run, local public enterprises have a competitive cost advantage over private suppliers²⁰.

Managers and local politicians often point out that local public enterprises provide special management skills and technological know-how for local administration, which is said to be more costly, or even impossible, to buy from private suppliers. This might be correct if the governance costs of outsourcing services to private firms were prohibitively high. But the services in question, for example IT-services, internal consulting, and so on are usually provided in competitive markets, and therefore, in-house provision in public enterprises may increase the crowding-out of private suppliers as well.

German local public savings banks (“Sparkassen”) are heavily involved in sponsoring activities in their market area. They support sporting events and co-finance cultural, scientific and social projects as well as job training measures²¹. Proponents argue that local sponsoring would improve “soft” factors of location (image, cultural environment, social infrastructure) or even increase the local stock of human capital as a “hard” factor of location for private enterprises. But these supposed positive externalities are often hard to quantify. The benefit-cost ratio in particular might be overestimated. Furthermore, it is difficult to separate altruistic gifts from systematic and tax-deductible methods of advertising or public relations²². Finally, it is worth debating whether local sponsoring activities lead to crowding-out of private suppliers (for example by providing computer courses for seniors) or prevent interest rates for saving deposits from being raised, or fees from being reduced.

Local politicians often argue that liberalisation or privatisation of public production would destroy local jobs or move them out of the area. This might be realistic for central business administration (strategic planning, financial management, accounting, central management, legal affairs, public relations, laboratory services, and so on), but should not be expected for technical personnel (operating and maintenance personnel, customer services, meter maintenance, technical support, and so on). From a spatial allocation point of view, it should be discussed whether local public enterprises are actually the brake or the motor for regional economic development. On the one hand, they might induce or boost existing positive agglomeration effects (primarily urbanization effects). On the other hand, a large

²⁰ For further discussion of cross-subsidisation at local level, see for example Haug (2003).

²¹ For an overview of the sponsoring activities of German local savings banks see Wengler (2001).

²² Sponsoring activities in schools (“Schulsparen”, “Planspiel Börse” and so on) or gifts (marked with the “S”-logo) for students’ sports teams or other youth organisations seem to be mainly instruments for acquiring young customers.

local public sector might delay necessary structural changes by tying up resources that could be employed more efficiently otherwise.

Lobbyists often emphasise social – in other words, distributive – goals of local public enterprises that they consider more desirable from an ethical point of view than goals like maximising profits or increasing internal efficiency. If these goals were taken into account, evaluating performance would be more complex than simply calculating balance sheet indicators. This might further increase control costs and make it more difficult to set up proper incentive schemes for public managers.

As a consequence of their – declared – “social responsibility”, local public enterprises feel obliged to offer more favourable working conditions than their private counterparts. But restrictive regulations on payment and dismissal make it difficult for local public enterprises to attract qualified and motivated personnel or to dismiss unqualified and unproductive employees.

One of the most heatedly discussed aspects of the private provision of local public goods and services are the potential distribution problems: low-income households might be unable to afford water, sewage treatment, waste disposal, electricity, gas or heating after privatisation. Consumer unionists in the UK, for example, are currently complaining about rising electricity prices for low income customers²³. Apart from the fact that liberalised markets for other essential goods such as food and beverages in Western Europe have not caused starvation or other severe shortages since World War II, equity is often (knowingly?) misinterpreted. Instead of concentrating on the small minority of poor people who really cannot afford some essential goods at market prices, the public providers subsidise everybody through provision of water, gas, electricity, public transportation, cultural events and so on at prices often far below their costs. Public production and provision of these goods and services may only make sense if the governance costs of introducing a voucher system, such as the costs of assessing neediness, exceed efficiency losses from public production and provision.

Although many problems mentioned in the discussion here must remain unresolved, the theoretical and empirical findings provide some evidence of increasing welfare if market interference by local governments is reduced. The next section will deal with the question as to why necessary reforms in the field of public services of local governments are still being blocked.

²³ See Jewell (2003). The complaints have to be softened if one takes into account the higher risk of illiquidity of low-income households. Higher prices per kilowatt-hour for prepayment meters (which are preferred by or mainly offered to low-income households without checking accounts) may either reflect higher costs of this technology or include some kind of risk-surcharge. Prepayment meters are used as substitutes by the former public British electricity providers for the rather unpopular practise of disconnecting notoriously defaulting customers.

3. Explaining the Costs of Institutional Change
in the Field of Local Public Services

So far, in the public discussion of local public services, it has been a commonly used strategy to oppose to privatisation by quoting negative examples in some fields of local public services, where privatisation has led to increasing costs or losses in service quality. But such examples are not a conclusive argument against privatisation in general. Failures in privatisation may be due simply to an incorrect choice of institutions under which the privatised services are to be produced. Yet several groups are not even willing to *discuss* proper institutional arrangements, and oppose any kind of privatisation. Why are these groups acting as agents of preservation, and why have they been, up to now, so successful in preventing the agents of change from implementing a comprehensive reform in the field of local public services of general interest?

In general, the groups that benefit from the status quo can be identified as agents of preservation. In order to identify them, one must ask which actors are benefiting from the existence of local public production units. On the other hand, one may identify the relevant agents of change in the field of public policy by looking at those groups for which the current institutional setting is leading to losses of welfare.²⁴ Table 1 gives an overview of the groups that are supposed to act as agents of preservation or as agents of institutional change, respectively, in the field of local public “services of general interest”.

Table 1

The Incidence of Costs and Benefits from the Institutional Status Quo
in the Field of Local Public “Services of General Interest”

Groups affected by welfare losses from the institutional status quo	General Public EU-Commission (in general) Small Private Firms from Various Fields of Production (which could be ready to replace public production units)
Beneficiaries of the institutional status quo	Members of the Local Production Units Local Governments Trade Unions Charities Environmentalist Groups Small Private Firms which are Integrated in Local Pub- lic Production Relevant Number of Parliamentarians at the National Level and the EU Level National Governments

Source: Authors’ graph.

²⁴ For the general approach and the main hypotheses of the Economic Theory of Institutional Change, see Rosenfeld (1996), pp. 37–61, Rosenfeld (1997), and Rosenfeld (1999).

Section 2.2 described the different categories of welfare losses that result from the prevailing public production of services of general interest. These costs must be born mainly by the general public. Additional costs that arise exclusively for particular groups in society will be discussed later. With regard to the costs to the general public, one may assume that they are distributed more or less equally. The expected general increase in economic welfare from privatising public production may be interpreted as a pure public good. Therefore, in accordance with Mancur Olson's famous "Logic of Collective Action"²⁵, the probability that some agents would be willing to become involved in helping to reduce these costs is low compared to costs distributed more unequally. The probability is rather high for some groups or agents, with the result that they are willing to overcome the "free rider" problem. The EU Commission has, in the past, played a prominent role as an agent of change in the field of privatisation²⁶. This may be explained by the fact that the EU Commission is not elected by the European Parliament and has no direct links to the governments of the EU member countries; therefore, the EU Commission has enough leeway to articulate demands that are in accordance with general welfare. In addition, the EU Commissioners, like bureaucrats in general, want to expand their sphere of authority. And as liberalisation of markets is a general task of the EU Commission, it is obvious that a successful privatisation process may be claimed as a success of the EU Commissioners and may expand the commission's sphere of authority. But at the moment, the EU Commission seems to have less power than the agents of preservation.

A group with additional costs of the institutional status quo is that of the private enterprises that are *potential private suppliers of the services currently produced by local public units*. For this group, the monopoly of the local public production units prevents them from entering a market and from expanding their markets. Europe already has some very large private firms in the energy market, which was opened up to private firms in the 1990s, and in the water market, which is open to private competitors in some countries. But in most other fields of local public service production such as in the childcare services sector or in the street cleaning business, there are currently only small private firms. As it is rather costly to organise the interests of many small firms from various fields of production, it is not surprising that the interests of these firms are seldom articulated in the debate about privatising local public production. The general representatives of private firms, such as the chambers of commerce, sometimes try to criticise the status quo. However, as will be shown later on, some private firms are, at the moment, among the beneficiaries of local public production, and these firms may hinder chambers of commerce from taking severe actions against local public production. With regard to the water market, a few large companies exist in Europe. It seems realistic to assume that these companies would be interested in entering new markets. But

²⁵ Olson (1965).

²⁶ For an example, see Blankart (2002), pp. 1 – 7.

perhaps even the possible gains are not high enough to provide an incentive for promoting institutional change. The amount of profits depends on the country-specific regulations in the water market. The fact is that today, the profits in the water market are comparatively low. It may be that expanding into third-world countries is more attractive for the large companies in question than struggling for liberalisation within the EU. Thus, these firms prefer the exit option to the voice option²⁷. And of course, firms that do not yet exist due to the current lack of market-oriented activities in the field of local public services have no voice option within the political process at all.

The *agents of preservation* in the field of local public service production are better organised than the agents of change, because the incentives of several groups to preserve the status quo are stronger than the incentives for reform described above. Apart from the benefits from the institutional status quo, which have already been laid out in Section 2.3, there are also some categories of benefits to certain groups that are seldom mentioned in the public discussion. Apart from the members of the local public production units themselves, who obviously have no incentive to change the present state, and the associations of (local) public producers at the regional, federal and EU level²⁸, local government as a whole also benefits from the status quo. As has been described in Section 2.3, local public production units may be used as substitutes for local taxation (advantages of the so-called *Querverbund*). By subsidising some local public services with profits from other branches, local politicians can camouflage the real costs of the services in question. In addition, local policymakers may claim that local public production units secure jobs and the re-investment of profits within the region. The special knowledge of the members of the local public production units may be used at low cost for the purposes of the local public general administration, and the local public production units often sponsor social events that are in the interests of local policymakers.

Additional benefits may arise if local policymakers use public production units to compensate politicians for past favours. These politicians may get jobs in the management of the local public production units, or on supervisory boards. Only under extremely unfavourable conditions will local authorities agree to privatise a local public production unit. This may be illustrated by the example of the German city of Stralsund. In this case, the main reason for the city managers' plan to sell the local public savings bank to a private investor was that, according to the plans of the state government of Mecklenburg-Western Pomerania, the city of Stralsund would lose its present status as an "independent city" (*kreisfreie Stadt*) and all the

²⁷ See Kemper (2001), pp. 24–34.

²⁸ The "European Centre of Enterprises with Public Participation and of Enterprises of General Economic Interest (CEEPP)" with its German section "Gesellschaft für öffentliche Wirtschaft (GöW)" or the German "Verband kommunaler Unternehmen (VKU)" are examples of influential lobbying groups of public producers.

assets of its local public savings bank. Therefore, in order to secure the assets for the local economy, the city managers have brought up the plan to sell the bank.²⁹ A further factor could be that local governments and their associations fear a potential loss of authority because of the fact that privatised former local public monopolies may be easier to regulate in Brussels or Berlin than by local government.

All these arguments suggest that local policymakers currently face a lock-in problem. Although privatising public production could be advantageous for local public finance and for local economic growth (see Section 2.2), the benefits of maintaining the institutional status quo forces them to act as agents of preservation.

The *trade unions* are interested in maintaining the status quo because their main goal is to secure the jobs of those who currently have a job, for example the employees of local public production units. Public employees have more long-term income security, more rights than private employees, and more “leisure on the job”. This means that they have enough time to participate in and organize trade union activities, and thus extremely high incentives to preserve the current advantages of their public jobs. In addition, the trade unions may try to secure their influence on the local public production units, where the rights of co-determination in plants and the tendency to keep collective wage agreements are, in, general higher than in private firms.

Charities fear that private firms will raise the prices on several services that are important for people with lower income. But what they may fear most of all is that a general tendency towards liberalisation could force the charities to raise the quality of their own services and to work to improve service efficiency. In addition, higher-income earners may easily be won over by the agents of preservation as opponents of liberalisation. Nearly every household today enjoys some benefits from the local production units. Children attend public kindergartens, use public libraries, or take advantage of other publicly financed leisure activities; parents use local public transport to go to work; everybody consumes water from local public water utilities. For specific actions against privatising-plans within a community, it is comparatively easy to organise the consumers of these services to protest.

Small private firms that are integrated today in the local public production of services also tend to support the agents of preservation, because they do not know whether a private producer would also engage them.

Environmental groups pretend that the high standards of environmental protection (for example in the cases of water supply or sewage treatment) are only due to public production. Of course, many environmentalists know that this is not the case. But they fear that new regulations – which would be necessary for the private production units – could not go as far as the present practice.

²⁹ Refer to “Stralsund stoppt Verkauf der Sparkasse” (2004).

Many politicians at the federal level not only have their roots at the local level, but also have strong connections to local government, even when they have been elected into the national parliament. Table 2 illustrates this for the case of the German Bundestag. About 40% of the members of the Bundestag are connected to the local level. It may be expected that at least some of the politicians in question will try to support the case of local government at the federal level, at least as long as it does not cost them too much³⁰. No single member of parliament can gain much directly if he or she votes for liberalisation. The situation would be different if the federal government had a core interest in privatising local public production, but this is not the case. The national governments, like the local governments, fear losing authority to the EU. In addition, the national government, like the local policymakers, wants to protect national production units from international competitors.

Table 2

Members of the German Bundestag and their Relation to Local Public Organizations

(1) Members with Relevant Relations to Local Public Organizations ^a	28.7%
(2) Members with Strong Relations ^b to Local Public Organizations	11.3%

^a Persons who, during their present membership, are also members of local parliaments or have been, in the past, local public bureaucrats, leading local public authorities, or employees of local public production units; former activities in local parliaments are not included. – ^b Persons who, during their present membership, are also members of local public production units (mainly, directors of local savings banks (*Spar-kassenvorstände*), and so on) or members of federations of local units; a and b are independent from one another.

Source: Deutscher Bundestag (internet pages), authors' research and calculations, based on 247 members of the Bundestag.

Probably, at the EU level, a high proportion of parliamentarians would also have strong relations to local governments. It is likely that the importance of these relations is even greater on the EU level than on the national level. The reason for this is that members of the EU parliament have, at the moment, no obligations to the "European Government". Parliamentarians today are obliged only to their local or regional electorates, not to a certain general European policy agenda, and need not support an "EU government"). This may explain why the EU parliamentarians have, in recent years, worked as some of the most important agents of preservation in the field of local public production³¹. By opposing the EU commission, EU parliamentarians have the chance to increase their popularity at no cost to themselves.

³⁰ In the case of reforming the German system of local taxes, we have quite a different situation, because a reallocation of public funds to the localities will mean losses of money for the other levels of government.

³¹ Refer back to Section 1.

4. Conclusions

A comprehensive privatisation of local public services will only be achieved if (a) the general influence or power of the agents of preservation on the political outcome is reduced, (b) some of the agents of preservation could be compensated in other fields of public policies, and / or (c) the general theories on the outcome of privatisation are changed by more “good practices” in this field.

With regard to (c), it should be noted that private production does not *always* have to be the more efficient alternative. There is thus a need for more research into the question of how to arrange for the efficient production of publicly provided services. A necessary condition for such research would be improved access to the data of local public production units.

What other strategies can help overcome the reform blockade in Europe? Steps have to take place both in the member states and at the EU level. With regard to alternatives (a) and (b), and in view of our discussion in Section 3, it is not realistic to assume that new groups of agents of change could appear on the scene in the coming years. Furthermore, the currently existing groups of agents of change will probably not be able to increase their political power. Therefore, the present lock-in problem will only be solved if it is possible to reduce the benefits of the institutional status quo. As a consequence, at least some of the agents of preservation could be brought in to give up their resistance to privatisation.

A first step toward a solution at the national level would be a comprehensive reform of local public finance in general. If the local units did not depend on revenues from local public services, their benefits from the institutional status quo would be reduced. As local politicians and local public bureaucrats have installed local public production units with the aim of maximising their budgets and their authority, a “functional reform” could compensate the local agents for losses in the public production sector. One proposition for expanding the current local public authority is, for example in Germany, the decentralisation of responsibility for police forces from the Länder to the local level.

Another step could focus on the prices of local public services. The cost-recovery rate of most locally provided services could be improved. In the long run, all local public goods and services that are characterised by rivalry in consumption, and excludability, ought to be supplied at cost-covering prices. If politicians were planning to tolerate future deficits, for example in public transportation, for political reasons, losses should be covered by local tax revenues rather than by cross-subsidising services with deficits by profitable services. Thus, profits in some branches of the local public enterprise sector would lose value for local governments as a source of revenue. This might reduce the opposition of local politicians to liberalising the local public sector.

As a proper instrument for increasing equity, a voucher system (for example, the “Hamburg Model” for childcare services) could be introduced at the local level.

The obligatory voucher system could be applied to privatised, formerly local public goods and services or to local public goods and services offered at cost-covering prices. Vouchers would be issued to low-income households, which would grant them access to certain goods and services. This may weaken the arguments of supposed supporters of the “poor”, especially of charities. It has already been mentioned in Section 2.3 that the present system of subsidising all customers by supplying goods at non-cost covering prices could be retained only in cases where transaction costs of the voucher system plus costs of regulating the relevant private sector would be higher than efficiency losses of public production.

With reforms in the same direction, the benefits of the status quo claimed by environmental groups could be reduced. To improve the environmental quality, more direct measures could be implemented (for example higher standards for reducing air pollution by cars) so that local public production (for example of local public transportation services) would lose its relevance.

Furthermore, some institutional reforms at the EU level should be undertaken: The Parliament of the European Union has turned out to be the main obstacle in realising liberalisation proposals for the last few years. The recent excessive resistance to the commission’s policy might have resulted from the parliament’s awareness of its powerlessness. Therefore, the European parliament ought to be transformed into a real parliament representing a real “division of powers”, including budgetary rights, the right to elect the “executive” of the European Union (the commission), and so on.

References

- Armstrong, T. O./Leppel, K.* (1996): “Will Deregulating Water Utilities Result in Natural Monopolies?”, in *The Southern Business & Economic Journal*, 20 (1), 44–49.
- Baumol, W. J./Bailey, E. E., Willig, R. D.* (1977): “Weak Invisible Hand Theorems on the Sustainability of Multiproduct Natural Monopoly”, in *American Economic Review*, 67 (3), 350–365.
- Beitia, A.* (1995): “Quality Regulation in the Monopoly Case: the Impact of Information” in *Revista Española de Economía, Monográfico “Regulación”*, 41–60.
- Blankart, C. B.* (2002): “Modelle der Daseinsvorsorge aus EU-rechtlicher und ökonomischer Sicht”, in *WuW*.
- Bongaerts, J.* (1982): “Marktzutrittsresistenz eines natürlichen Mehrproduktmonopols”, in *Regulierung öffentlicher Unternehmen*, edited by C. B. Blankart, M. Faber, Königstein/Ts., 39–49.
- Haug, P.* (2003): “Quersubventionen in kommunalen Unternehmen: Praxis und theoretische Wertung?”, in *List Forum für Wirtschafts- und Finanzpolitik*, 29 (3), 197–215.
- Jewell, K.* (2003): “Manipulated, Misled, Ignored, Abused: Residential Consumer Experience with Electric Deregulation”, report of a project sponsored by PSRIU (University of Green-

- wich, UK) and Consumers Union (New York), <http://econwpa.wustl.edu/eps/pe/papers/0401/0401005.pdf>, read March 3, 2003.
- Kemper, B. (2001): "Einführung – Chancen und Herausforderungen für die deutsche Wasserwirtschaft", in *Wasserwirtschaft im Aufbruch*, edited by E. Büscher, Cologne, 19–38.
- Kim, H. Y. (1985): "Economies of Scale in Multi-product Firms: an Empirical Analysis", *Economica*, 54, 185–206.
- Laffont, J. J./Tirole, J. (1991): "Provision of Quality and Power on Incentive Schemes in Regulated Industries", in *Equilibrium Theory and Applications*, edited by W. A. Barnett, Cambridge.
- Megginson, W. L./Netter, J. M. (2001): "From State to Market: A Survey of Empirical Studies on Privatization", *Journal of Economic Literature*, 39, 321–389.
- Mueller, D. C. (2003): "Public Choice III", Cambridge.
- Mühlenkamp, H. (1999): "Eine ökonomische Analyse ausgewählter institutioneller Arrangements zur Erfüllung öffentlicher Aufgaben", Baden-Baden.
- Neal, K. et al. (1996): "Restructuring America's Water Industry: Comparing Investor-Owned and Government Water Systems", Reason Public Policy Institute, Policy Study No. 200, <http://www.rppi.org/ps200.html>, read September 2, 2003.
- Niskanen, W. A. (1971): "Bureaucracy and Representative Government", Chicago.
- (1978): "Competition among Government Bureaus", New York.
- Noll, R. G. (1989): "The Politics of Regulation", *Handbook of Industrial Organization*, Vol. II, edited by R. Schmalensee, R. D. Willig, North-Holland, 1253–1287.
- Olson, M. Jr. (1965): "The Logic of Collective Action", Cambridge, Mass.
- Rödl & Partner (2003): "Effizienz- und Qualitätsuntersuchung der kommunalen Wasserversorgung in Bayern (EffWB), Abschlussbericht für das Erhebungsjahr 2000", Nuremberg.
- Rosenfeld, M. T. W. (1996): "Evolution öffentlicher Aufgaben und Ökonomische Theorie des Institutionellen Wandels", *Volkswirtschaftliche Schriften*, 466, Berlin.
- (1997): "The Economic Theory of Institutional Change (ETIC) as an Approach to Explain the Emergence of Public Activities", *Zeitschrift für Wirtschafts- und Sozialwissenschaften (ZWS)*, 117, 291–312.
- (1999): "Wo stehen wir mit dem kooperativen Föderalismus in Deutschland?", lecture at the 62nd ARGE meeting, Beihefte der Konjunkturpolitik, 49, 2000, 53–99.
- Sheshinski, E. (1976): "Price, Quality and Quantity Regulation in Monopoly Situation", in *Economica*, 43, 127–37.
- "Stralsund stoppt Verkauf der Sparkasse" (2004), *Frankfurter Allgemeine Zeitung*, 53, March 3, 2004, 13.
- Wengler, M. (2001): "Gemeinnütziges Engagement der Sparkassen: Umfang und regionale Wirkungen", *Wirtschaft im Wandel*, 12, 298–305.
- Williamson, O. E. (1985): "The Economic Institutions of Capitalism", New York.

The Reform of Local Public Services of General Interest in Europe

Comment

By Charles B. Blankart*

1. The Paradox of Deregulation

Economists often face of the following kind of problem: They make a profound economic analysis and then come to the conclusion that the present situation could be substantially improved by an institutional reform. They publish their proposal and find support in scholarly discussions. But when proposing it to politicians for political action, opposition emerges and brings the intended reform to a halt. It seems to be a paradox that what is considered as good from a welfare economic point of view turns out to be bad from a political point of view.

Martin T.W. Rosenfeld and Peter Haug intend to explain this *paradox*. Economists have had some success in promoting deregulation, as the example of telecommunications deregulation demonstrates. But the initial stream of political action has now dwindled to a few drops in a dry watercourse.

Turning to local public services, the authors discuss the arguments in favor of deregulation – mainly from the standpoint of the economics of information and of capital markets – and then turn to the arguments of those who are against deregulation, among which those of the so called “instrumentalists” who praise the gross benefits generated by public and communally oriented enterprises. So local public enterprises are said to promote the local infrastructure, to support local public transport, to generate know-how in IT services, to guarantee “affordable” consumer prices, and to generate a soft climate for attracting investors, etc. But these authors notoriously neglect the costs of such policies: the financial burden plus the excess burden to finance the cross subsidization. When taking account of both, the balance is almost always negative.

Regulation, to conclude, though often inefficient, is nevertheless an arrangement that seems to be institutionally stable. How can this be explained? Rosenfeld and Haug suppose that local interest groups and locally based politicians are against

* Humboldt-University Berlin

deregulation and privatization. They block deregulation and privatization because they benefit from it. The authors argue:

“A comprehensive privatisation of local public services will only be achieved if (a) the general influence or power of the agents of preservation on the political outcome is reduced, (b) some of the agents of preservation could be compensated in other fields of public policies, and / or (c) the general theories on the outcome of privatisation are changed by more ‘good practices’ in this field.”

The authors’ conclusion, however, raises an even deeper paradox: Why does this not happen? In an open society all inefficiencies should be eradicated sooner or later. How can it be that such an apparently inefficient arrangement persists over a long time? I believe that the theory of expected vote maximization may provide some insights into this paradox (Section 2). These insights can be illustrated by the recent trends in regulation of the services of general economic interest on the EU level (Section 3). Eventually globalization may eradicate the existing regulatory system (Section 4).

2. Interest-Group Influence in the Expected Vote Maximization Approach

The logic of the model of expected vote maximization is the following: If, under two party competition, the candidates cannot exactly predict voters’ reactions to policy changes, they maximize the probability π of electoral support. More specifically, they distribute a given sum of money Y , presumably obtained from public enterprises, in such a way among the voters that their expected votes EV are maximized (see D. C. Mueller 2003, 252 – 254).

$$(1) \quad EV_1 = \sum_{i=1}^n \pi_{1i} = \sum_{i=1}^n f_i(U_i(y_{1i}) - (U_i(y_{2i}))) + \lambda(Y - \sum_{i=1}^n y_{1i})$$

where U_i is the utility of voter i and y_i is the transfer obtained by the individual i . Candidate 2 chooses the complementary strategy $1 - EV_1$. Under the assumption that $f(\cdot)$ and $U(\cdot)$ are continuous and concave, both candidates will choose the same stable platform satisfying the following first order conditions:

$$(2) \quad f'_i U'_i = \lambda = f'_j U'_j .$$

The f ’s are the probabilistic responses to the utility differences promised in the candidates’ platforms. If these responses are the same for all individuals, equation (2) can be transformed to equation (3)

$$U'_i = U'_j \quad \text{for all } i, j = 1, n$$

which can be interpreted as the outcome of a Benthamite social welfare function

$$(4) \quad W = U_1 + U_2 + U_3 + \dots U_i + \dots U_n .$$

Conclusions:

1. Expected vote maximization generates endogenously a stable equilibrium with attractive welfare economic properties.
2. The equilibrium is characterized by differentiated transfers.
3. Regulated public enterprises are ideal tools for imposing differential burdens and channelling differential transfers to selected voter groups. For they are not subject to as strict formal regulations as taxes.
4. Deregulation is undesirable by politicians because it would squeeze the rents available to politicians.

3. The Role of the European Institutions

What has been shown in theory can be illustrated in practical politics. A competitive regime for services of general economic interest is difficult to enforce in the long run. The destiny of Article 86 EC may to illustrate the hypothesis.

In 1957 the Heads of States signed the Treaty of Rome in which Article 86 contains the regime of competition that should be applied to the services of general economic interest¹: “Undertakings entrusted with the operation of services of general economic interest or having the character of a revenue-producing monopoly shall be subject to the rules contained in this Treaty, *in particular to the rules on competition*, insofar as the application of such rules *does not obstruct* the performance, in law or in fact, of the particular tasks assigned to them” (par. 2, emphasis added). Par. 3: “The Commission shall ensure the application of the provisions of this Article and shall, where necessary, address appropriate directives or decisions to Member States.”

From the wording of these paragraphs one can conclude that, in general, competition rules should be applied. Only where the application of these rules does obstruct the performance of the service, exceptions to the rules of the Treaty could be considered. The burden of proof should rest on those who apply for an exception. Consequently, no further decision-making was foreseen in the Treaty, and the management tasks have been handed over to the Commission (par. 3).

This clear-cut regime has, however, been eroded continuously in the subsequent years by the interference of the national governments. In 1995, the European Centre of Enterprises with Public Participation and of Enterprises of General Economic Interest (CEEP), a powerful interest group on the European level, postulated that

¹ This regime should apply to local public services too which, today, are often provided or could be provided by large cross border networks affecting intercommunity trade.

obligations and rights of public enterprises in the EU should be given back to national responsibility and sovereignty.

The national governments supported such pressures and added Art. 16 EC to the Treaty in 1997 which declared the services of general economic interest a domain of the Union *and* its Member States. " ... The Community and the Member States, each within their respective powers and within the scope of the application of this Treaty, shall take care that such services operate on the basis of the principles and conditions which enable them to fulfil their missions."

National responsibilities for services of general economic interest have been addressed even more distinctly in Art. II – 36 of the Charter of human rights in the year 2000. "The Union recognises and respects access to services of general economic interest as provided for in national laws and practices, in accordance with the Constitution, in order to promote the social and territorial cohesion of the Union." This article implies a clear departure from the application of Union wide competition laws notwithstanding that Art. 86 of the Treaty (or Art. III – 6 Draft EU Constitution, see below) should still be maintained.

The new European Constitution has made a last step towards a re-nationalization of the services of general economic interest by adding to the former Art. 16 of the Treaty that: "European laws shall define these principles and conditions." (Art. III – 6 Draft EU Constitution), meaning that from now on the Council and no more the Commission shall be responsible for the regulation of the services of general economic interest. In other words: What has been decided to be a basic principle of the Union in 1957 (and hence allocated to Commission for execution) shall be discussed anew under the new Constitution.

To conclude, the Treaty of the European Union was not a strong enough institution to overcome the endeavor of national and local politicians to re-nationalize public enterprises and to use them as means for redistributing rents in political competition for votes. Political competition overwhelmed over economic competition.

4. Globalization and Systems Competition

Does that mean that the laws of economics are weaker than the laws of politics? Not necessarily. A challenge to politics comes from globalization. The political equilibrium with some paying for and some benefiting from cross subsidization may one day no longer be sustainable. Globalization enforces the law of one price, and price discrimination breaks down.

Of course, politicians may try to resist to globalization. But the price of resisting competition will continuously increase: the price is declining growth rates. For Germany, this price is already substantial. The cumulative growth rates from 1992–2004 are 33 percent for the OECD total whereas they are only 17 percent for

Germany. The difference is 16 percentage points. Hence the growth of the OECD total is nearly twice as large as that in Germany. Eventually, the price of the growth foregone may be so high that politicians will find it beneficial to switch to growth policy in order to gain more votes.

References

Mueller, D. C. (2003): "Public Choice III", Cambridge University Press, Cambridge.

Social Welfare Reform and the Low-Wage Labor Market in Germany: What Works and What Doesn't?

By Viktor Steiner*

Abstract

Various social and labor market reforms, such as the reduction of unemployment benefits and social assistance, the substitution of these transfers by “in-work” benefits, and stricter work requirements, have been suggested in the German economic policy debate with the aim to increase employment in the low-wage sector thereby reducing unemployment among low-skilled workers. This paper shows that, although pointing into the right direction, these reforms as implemented by the German government’s “Agenda 2010” will not have the intended impact on the labor market because they do not properly address the incentive problems inherent in the German welfare state. I propose a coherent social welfare reform which would increase work incentives for unemployed people with low earnings potential. It is shown that the proposed reform would lead to a marked increase in employment in the low-wage sector of the German economy without resulting in an unsustainable increase in social expenditures.

JEL classification: H24, H31, I38, J22

Keywords: Social Welfare Reform, Work Incentives, Unemployment

Among economists and the general public alike, the view has increasingly become popular that relatively generous income support programs for unemployed people have contributed to the high level of unemployment in Germany.¹ There are basically two arguments for this belief. The first is that, due to the small difference between net income from full-time employment in a low-wage job and the level of social assistance for certain household types, there is no or very little financial incentive to take up work in the regular economy. In other words, the

* Corresponding author: Free University Berlin and DIW Berlin. Address: DIW Berlin, Königin-Luise-Str. 5, 14195 Berlin, E-mail: vsteiner@diw.de.

I thank the German Science Foundation (DFG) for funding under the project “Arbeitsmarktflexibilisierung durch Subventionierung der Sozialbeiträge im Niedriglohnbereich – Eine Mikrosimulationsstudie für Deutschland”.

¹ Recent prominent subscribers to this view include Sinn et al. (2002), the German Council of Economic Advisors (Sachverständigenrat 2002), the Scientific Council of the (former) Federal Economics Ministry (Wissenschaftlicher Beirat 2002), and the European Economic Advisory Group (EEAG 2003).

benefit withdrawal rate for social welfare recipients is simply too high, and this negative incentive effect is reinforced by high social security contributions and income taxes which already set in at relatively low earnings. The other argument refers to the role of social welfare in sustaining a relatively high market wage: the social minimum effectively sets a lower wage floor for low-skilled labor. Although the basic problem according to both views is really the relatively high level of social welfare, the first view stresses the supply side, whereas the second emphasizes the demand side of the market for low-skilled labor. Of course, both sides of the labor market are affected, and these arguments are complementary rather than substitutes for each other.

Various approaches have been suggested in the literature and in the economic policy debate to make work financially more attractive relative to non-work. At the one extreme, there are policies which “wield the stick” of time-limited welfare entitlement and strict work requirements. These instruments have played a major role in the US welfare reforms of the 1990s (Blank 2002, Moffitt 2003), but also in some European countries, like Denmark, Sweden and the United Kingdom (see EEAG 2003). At the other extreme, there are reforms mainly relying on the “carrot” of earnings-related subsidies for people with low earnings potential. As discussed below, this may be administered by reducing the benefit withdrawal rate of social welfare, subsidies to social security contributions, a negative income tax on low earnings, or a combination of these instruments. The Earned Income Tax Credit in the United States and the Working Family Tax Credit in the United Kingdom are well-known examples of earnings-related subsidies. There is also an intermediate approach which combines the carrot of earnings subsidies with the stick of a reduction or even a complete cut of social welfare for those who choose not to work. The US welfare reform of the 1990s and the “New Deal” in the UK are recent examples of such reforms (see, e.g., Blank 2002, Blundell 2000).

In Germany, too, there have recently been various proposals to reform social welfare with the intention to increase work incentives in Germany, thereby reducing unemployment of low-skilled workers and increasing employment in low-wage labor markets. One of the most controversial issues in the recent German debate relates to the alleged disincentive effects of social assistance and unemployment assistance. These programs, which constitute the German version of “social welfare”, provide means-tested income support at a socially defined minimum income without a pre-specified time limit. There has recently also been some discussion on time-limits of social welfare and the enforcement of strict work tests in Germany, and several communities have been experimenting with time limits and strict work requirements (see Feist and Schöb 1998, Sinn et al. 2002).

The recent reform steps of the German government under its so-called “Agenda 2010”, such as subsidized social security contributions on low earnings (“Mini-Jobs”), and the proposed integration of unemployment assistance and social assistance, are intended to increase employment in the low-wage sector of the economy.

However, as I will argue in this paper, the government's Agenda 2010 as implemented so far is not likely to achieve its intended aims. With the exception of the reduction of entitlement periods for unemployment benefits, the government's reforms seem unlikely to have a strong impact on unemployment in Germany. The reasons for this pessimistic view are, first, that the social welfare and the mini-job reforms do not set proper work incentives and, second, there is no effective work test reinforcing the potentially positive effects of these reforms.

To set the scene for the discussion of social welfare reforms in Germany, the next section presents some theoretical background on the relationship between earnings-related subsidies, work incentives and employment in low-wage labor markets. The focus is on the financial incentives and various labor supply responses induced by the introduction of an earnings-related subsidy on top of social assistance. In Section 3, I will briefly summarize and evaluate the main reforms under the German government's Agenda 2010 aimed at reducing unemployment and increasing employment in the low-wage sector of the economy. These include the reform of unemployment compensation and social assistance, stricter work requirements, and the recent mini-jobs reform. In Section 4, I will present a more radical welfare reform proposal for Germany and present empirical results on its expected work incentive, labor supply and employment effects. The concluding Section 5 summarizes implications for economic policy derived from the preceding analysis.

1. Earnings-Related Subsidies, Work Incentives, and Employment in Low-Wage Labor Markets

In the German debate on social welfare reforms, earnings-related subsidies have been suggested to improve work incentives and as a means to increase employment in low-wage labor markets (for recent summaries see, e.g., Buslei and Steiner 2003, Zimmermann 2003). Proponents of earnings-related subsidies expect them to improve financial incentives to take up low-wage jobs. To some extent, these proposals have been motivated by the experience with the *Earned Income Tax Credit* (EITC) in the United States². The EITC is a subsidy (a negative tax) on low earnings which depends on the level of earnings and on the number of children. In the year 2001, the subsidy rate for a family with two children and earnings below 10,020 US\$ was 40%, the maximum subsidies thus amounted to about 4,000 US\$; for earnings between 10,020 and 13,090 this subsidy remains constant. For earnings exceeding this latter amount, the subsidy is withdrawn at a rate of about 21%. This so-called phasing-out region ends at earnings of 32,121 US\$ when the subsidy has been completely withdrawn (see Moffitt 2003: 132).

² A similar earnings subsidy called the Working Families Tax Credit (WFTC) has recently also been introduced in the UK (see Blundell 2000, Blundell et al. 2000).

With about 55 million people supported by the EITC at a total cost of US\$ 26 billion in 2000, the EITC is now the largest social program in the United States. The large number of supported people is often considered *prima facie* evidence for the effectiveness of EITC type earnings subsidies. However, theoretical and empirical analyses do not necessarily support this view, as discussed below.

1.1 Social Assistance and Earnings-Related Subsidies

Economic analysis of the effects of social assistance and earnings-related subsidies on work incentives generalizes the household's budget constraint of the standard neoclassical leisure-income choice model by explicitly introducing these transfers into the model (see, e.g., Moffitt 2002). Figure 1 illustrates the basic

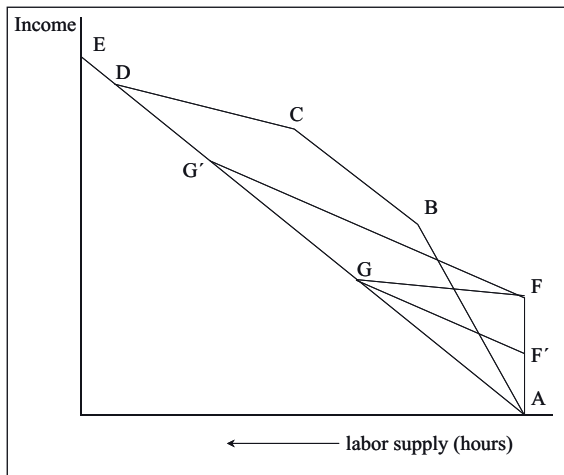


Figure 1: Social Assistance and Earnings-Related Subsidies

arguments. Assuming for simplicity that there are no other taxes or transfers on gross earnings, the budget constraint of the household in the absence of social assistance and the earnings-related subsidy is given by the line AGE. The budget segment ABDC represents an EITC type earnings subsidy with the three ranges mentioned above, with the segment CD representing the phase-out range. The budget segment AFG represents a social assistance program with a relatively high social minimum and a very high social assistance withdrawal rate of, say, 90%. Assuming that social assistance is not included in the calculation of the earnings subsidy, the total marginal tax (withdrawal) rate relevant for an individual's labor supply decision is the sum of the rates of the two programs. Since the earnings subsidy implies a negative tax rate at low levels of earnings, an individual entitled to social assistance

and working at the first phase of the EITC faces a low marginal tax rate because the benefit withdrawal rate is reduced by the amount of the EITC rate. In case of a 90% social assistance withdrawal rate and a wage subsidy on low earnings of, say, 40%, the overall withdrawal rate or marginal tax rate (negative tax and social assistance withdrawal rate) would be 50%.

To increase work incentives it is often proposed to reduce the social assistance withdrawal rate. As illustrated in Figure 1, however, a substantial reduction of the withdrawal rate is only feasible if the social minimum is simultaneously reduced substantially (from F to F' , say). Otherwise, the transfer range would extend well up into the phase-out region of the earnings-related subsidy, i.e. to point G' in Figure 1 (the lines $F'G$, representing a withdrawal rate of, say, 50%, and FG' are drawn to be parallel). This would not only be very expensive in budgetary terms, but would also imply a relatively high cumulative marginal tax rate on earnings in the phase-out range. Assuming a withdrawal rate of 20%, say, on the earnings-related subsidy in this range and a 50% social assistance withdrawal rate, the cumulative marginal tax rate on earnings below G' would be 70%.

These calculations assume that social assistance is not taken into account in the calculation of the earnings-related subsidy. If this was not the case, there would be little financial incentives to take up work for someone eligible to social assistance, depending on the respective withdrawal rate. If it approached the value of 1, as it is the case in Germany, entitled individuals would have no financial incentives at all to expand working hours.

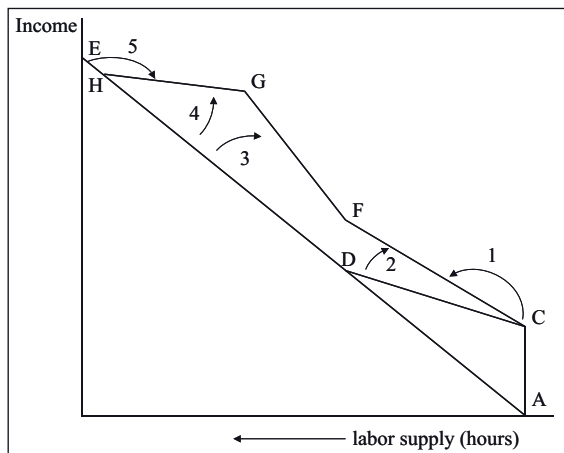
1.2 Work Incentives and Labor Supply Effects

The complex non-linearities in households' budget constraints introduced by the existence of means-tested social assistance and earnings-related subsidies induce a variety of labor supply effects typically ignored in popular discussions on the labor market effects of social reforms. Some of these effects are illustrated in the following figure (see, e.g., Moffitt 2002 for a more detailed discussion).

Similarly to the previous subsection, the budget segment ACD represents the social assistance scheme, where the respective benefit withdrawal rate determines the slope CD . The budget constraint $ACFGH$ represents the combined effects of social assistance and an EITC of the type defined in the previous subsection. As drawn in the figure, social assistance is assumed to have been completely withdrawn when the end of the phase-in range of the earnings subsidy (represented by the segment CF) is reached. Hence, it is assumed that at low earnings (within the phase-in range) social assistance is not taken into account in the calculation of the earnings subsidy. Net household income within the phase-in region therefore increases by a factor of $(1 - \text{cumulated tax rate})$, where the cumulated tax rate has been defined above. If social assistance was completely withdrawn in case of positive earnings, the budget segment CF would coincide with the line CD .

The labor supply responses induced by the introduction of an EITC on top of a social assistance scheme are illustrated in Figure 2. The arrows indicate that, depending on an individual's initial position on the budget constraint, the effect of introducing an EITC may go in either direction. For individuals initially entitled to social assistance and not working at all, and thus located at point C, the labor supply effect is unambiguously positive (arrow 1). For individuals located somewhere along the segment CD the effect could be either positive or negative. If the negative income effect dominates the positive substitution effect the individual will reduce her supply of working hours (arrow 2). For individuals already working beyond the phase-in range of the EITC the labor supply effect is also ambiguous (arrows 3 and 4). It will also depend on an individual's initial location on the budget constraint before the policy change. Finally, people already working beyond the phase-out range of the EITC may reduce their working hours to become eligible for the subsidy (arrow 5).

Figure 2: Work Incentives and Labor Supply



The net effect of the introduction of an earnings-related subsidy in addition to a means-tested social assistance scheme is therefore ambiguous in sign. It will depend on the parameters of the two programs, the distribution of the population along the different ranges of the EITC, and households preferences for income and leisure.

The analysis of labor supply effects induced by the introduction of an earnings subsidy or changes in social assistance rules becomes even more complex if labor supply decisions in couples households are determined as a joint decision. In addition to the effects mentioned above, an earnings subsidy affects the division of

spouses' labor supply in such a way to take the greatest advantage of the subsidy. This effect depends on household preferences and the detailed structure of the tax-benefit system (see Steiner and Jacobebbinghaus 2003, Steiner and Wrohlich 2004).

An important implication of the described incentive and labor supply effects of earnings-related subsidies and social assistance schemes is that any change benefiting the currently unemployed will also indirectly affect people currently not entitled to the program and induce them to change their behavior in such a way to benefit from it. This, in turn, will imply increasing costs of the program which have to be financed by higher taxes, thereby leading to further labor supply distortions. Because of these effects, the EITC and similar programs aimed at improving work incentives may be inefficient. Their desirability depends on the relative weight allocated to the various groups in calculating social welfare (Moffitt 2002, Saez 2002, Homburg 2003).

There seems also to be little reason for being too optimistic about the positive policy issue of the size of the labor supply effect of earnings-related subsidies. According to various empirical studies the work incentive and labor supply effects of these programs seem to be rather limited, partly due to the compensating effects discussed above. Some of these studies indicate that the expansion of the EITC since the mid-1980s has increased participation rates and hours worked of single parents, whereas labor supply of married women seems to have been reduced; the net effect on total working hours has been estimated to be positive but rather small (Eissa and Hoynes 1998, Meyer 2002, Hotz and Scholz 2003). Similarly, the introduction of the WFTC in the UK is estimated to have slightly increased labor force participation of single mothers but reduced employment of married women. Overall, the net effect of the WFTC on labor supply has been estimated to be a modest 30 thousand persons (Blundell et al. 2000).

1.3 Wage Adjustment and Employment

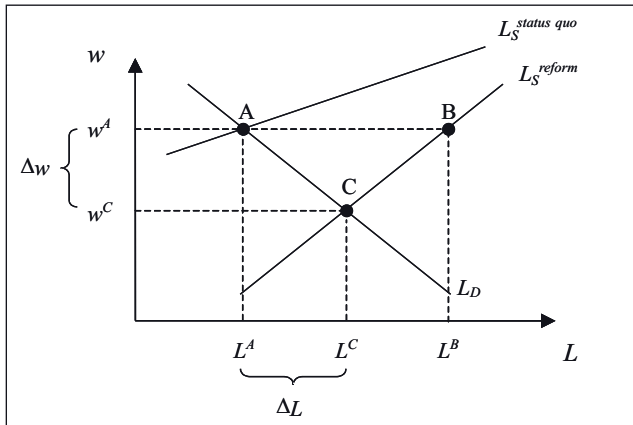
In the previous section, as in most of the discussion on potential labor market effects of social reforms, it has implicitly been assumed that the additional supply of labor will be matched by an increase in the demand for labor. Given the relatively small estimated total effects of the EITC and the WFTC, this might be a reasonable approximation. However, if a welfare reform does result in a substantial increase in the supply for labor, real wages have to fall so that labor demand increases sufficiently.

This is illustrated in Figure 3, where the positively sloped labor supply curve is drawn under the assumption that the social welfare reform increases the supply for labor at any given gross real wage. As drawn in the figure, this effect is particularly strong for those with a relatively low earnings potential. If the labor demand curve was perfectly elastic, market wages would stay constant at w^A and the em-

ployment effect would be given by the increase in labor supply, i.e. the distance AB.

However, allowing for a negatively sloped demand curve for labor, the increase in the supply of labor will only be absorbed by increased demand if the market wage falls. As illustrated in Figure 3, in the new equilibrium given by point C, the gross wage has fallen to w^C and employment has increased less than in B due to the adjustment of labor supply to the wage reduction. For a given shift of the supply of labor, the required wage reduction is determined by the size of the wage elasticity of labor demand: the larger (in absolute value) the labor demand elasticity, the larger the employment effect.

Figure 3: Labor Supply, Employment and Wage Adjustment in the Low-Wage Labor Market



Empirical estimates of own-wage elasticities of the demand for unskilled labor as reported in the literature vary substantially, but a value of -1 can be considered an upper-bound (see, e.g., Fuchs, Krueger and Poterba 1999, Franz 2003, chapter 4.4.2). My own estimates for Germany are somewhat below this value, and also differ by gender and whether they refer to total hours or persons (Steiner and Jacobebbinghaus 2003, Appendix 3, Table A2). However, if labor supply is rather inelastic, differences in the wage elasticity of labor demand will have little impact on the employment effect of the social reform anyway (see Figure 3).

Empirical estimates of (uncompensated) wage elasticities of labor supply as reported in the literature also vary greatly (Fuchs, Krueger and Poterba 1998, Blundell and MaCurdy 1999). Except for married women, for whom the within-period labor supply wage elasticity has been estimated to lie between 0.2 and 0.5 in most studies, labor supply seems to be rather inelastic in most countries. For Germany, my own estimates show that wage elasticities with respect to hours worked range

between 0.40 for west German wives to 0.10 for husbands, and are somewhat smaller for singles. Similar differences are also observed with respect to participation elasticities (see Steiner and Jacobebbinghaus 2003 Appendix 3, Table A2, Haan 2004).

These relatively small labor supply elasticities imply that financial incentives to take up work have to be substantial in order for a social reform to have a great effect on labor supply and employment in low-wage labor markets.

2. The “Agenda 2010” – Will It Work?

Shortly before the last election in 2002, the so-called *Hartz Commission* on behalf of the then incumbent German government presented various reform proposals aimed at “cutting unemployment by half” (Hartz Kommission 2002). Some of these proposals for a fundamental reform of the German labor market were taken up by the so-called “Agenda 2010” of the re-elected government. In the following, I briefly present and critically evaluate the most important of these reforms which aim at improving work incentives and increasing employment in the low-wage labor market.³

2.1 Unemployment Compensation and Social Assistance Reform

The government’s Agenda 2010 contains several reform proposals concerning unemployment compensation and social assistance which have become law in the meantime. Before this reform, unemployment compensation in Germany consisted of the unemployment benefit and unemployment assistance. Whereas the unemployment benefit is insurance-related and intended to replace previous earnings, unemployment assistance is tax-financed and means-tested, but nevertheless related to previous earnings. The means test applied for unemployment assistance under status quo regulations is also less strict than for social assistance which constitutes the basic safety net of the German welfare state (for a more detailed description see, Steiner and Jacobebbinghaus 2003, Appendix 1). Whereas unemployment benefits are time-limited, there is, in principle, no time limit for unemployment assistance and social assistance as long as the means test is passed and the formal work requirement (see section 3.2 below) is met.

Entitlement periods for unemployment benefits currently vary between 12 and 32 months for people aged above 57 years. An important reform under Agenda 2010 concerns the shortening of the unemployment-benefit entitlement period to

³ A critical evaluation of other reforms of the Agenda 2010 less directly relevant to the topic of this paper is presented by Steiner (2004); these reforms include organizational reforms of the Federal Employment Agency, employment protection legislation, subsidies for setting up one-person firms (“Ich-AG”), and the regulation of the crafts.

12 months generally and to 18 months for all unemployed persons over 55 starting from 2006. In the medium term, this will lead to a marked reduction of long-term unemployment, as long benefit-entitlement periods are a major reason for the high level of long-term unemployment in Germany, especially among older workers (see Steiner 1997, 2003).

Another important reform under Agenda 2010 concerns the integration of unemployment assistance and social assistance. The entitlement to unemployment assistance for an unlimited period after the exhaustion of unemployment benefit is another important factor for the high level of long-term unemployment in Germany (Steiner 1997, 2003). According to the planned reform, persons currently living on either unemployment assistance or social assistance but deemed able to work – everyone who can work at least three hours a day – will be entitled to a new benefit being referred to as “unemployment benefit II”. For those not able to work at least three hours a day, there will still be social assistance. The new “unemployment benefit II” contains a supplement to the current level of social assistance that takes account of the number of family members. In addition, recipients may obtain a wage subsidy for a limited period of up to two years if they take up a job. People entitled to this new benefit will also have access to advice and job placement through new *job centers* still to be established.

The unemployment assistance reform will reduce long-term unemployment among those who would have been entitled to it under current regulations. The reform is unlikely to strengthen work incentives for recipients of social assistance, however. Since the amount of the new unemployment benefit II is intentionally set to exceed the social assistance level, which is still means tested and almost completely withdrawn in case of own earnings, the already small gap between non-work and in-work net income shrinks even further, thus also reducing work incentives. In other words, the reform tends to increase the already extremely high benefit-withdrawal and marginal tax rates (in some cases considerably above 100%), and this will have negative effects on the number of hours worked for already employed people.

To improve work incentives for currently unemployed people the reform includes a wage subsidy granted to unemployed people who take up jobs in the low-wage labor market if considered “suitable” by the official in charge at the newly established job centers (see section 3.2). This subsidy should be limited to a period of two years. It will, therefore, only have permanent effects on unemployment if the productivity of those receiving this subsidy increased sufficiently over time to compensate for the withdrawal of the subsidy at the end of this two-year period. Whether or not this may happen seems doubtful, to say the least.

2.2 Work Requirements

The Agenda 2010 and the accompanying changes in the law also include some regulations concerning stricter work requirements for “employable” persons. The idea is that the introduction of a work requirement in terms of a minimum number of working hours will increase the effective supply of labor. The threat is that the new “unemployment benefit II” can be cut or be even completely withdrawn if the unemployed declines “suitable” job offers. In this case, the persons would fall back on social assistance provided the respective means test is passed. In principle, even social assistance could be cut if “suitable” job offers were declined⁴.

At the time of writing (March 2004), it is still not clear what “suitable” will eventually mean. There is currently some discussion that a job offer should only be deemed “suitable” if it comes with some sort of minimum wage. But even if the work requirement regulations remain as intended by the law, this is unlikely to change anything substantial regarding the effectiveness of the currently existing work requirements. On paper, these are already quite restrictive. However, in practice these work requirements have turned out to be rather ineffective because the labor office or social assistance unit has to prove the welfare recipient’s unwillingness to take up work. Although there are cases where communities have linked social assistance payments to a strict work requirement by way of a full-time public-works job (see, e.g., Feist and Schöb 1998, Sinn et al. 2002), this is still the exception rather than the rule.

Lack of job offers at the administration’s disposal is generally considered to be the main reason for the small number of temporary reductions or terminations of social welfare payments to recipients suspected not to be willing to take up work. The reason is that the administration fears to have to prove the case in court. One possibility would be to offer all recipients of social assistance who claim to be willing to take up but not able to find a regular job a public-works job at relatively unattractive terms. This, however, is not intended by the reforms of the Agenda 2010.

Instead the government hopes that more intensive job brokerage by the newly established *Personal Service Agencies* (PSA) may act as an effective work test for those drawing unemployment benefit II. These agencies are to be run privately on a subsidized basis or by the Federal Employment Service offices and are charged to give advice to unemployed persons and help them with their job-finding efforts. Initially, the German government expected the number of people attached to PSA to reach roughly 50,000 for 2003 as a whole. If this relatively number is not increased substantially PSA are unlikely to fulfil more than a marginal role in acting as a “work test”.

⁴ On the theoretical labor supply and welfare effects of work requirements see, e.g., Akerlof (1978), Besley and Coate (1992), and Homburg (2003), Moffitt (2002).

2.3 The “Mini-Jobs” Reform

In addition to the selective wage subsidy referred to in Section 3.1 above, the Agenda 2010 also proposed a subsidy of social security contributions (SSC) levied on so-called mini-jobs. Before the mini-jobs reform in April 2003, there was a lower earnings threshold of 325 €, up to which earnings were not subject to SSC on the side of the employee provided her weekly working time did not exceed 15 hours. The employer paid social security contributions of 22 percent of the employee's gross wage. Above the threshold of 325 €, the whole amount of earnings was subject to the normal rate of SSC, which is about 21 percent for employees and the same rate for employers. Earnings under the threshold were exempt from taxation in case the employee had no other income. Above that threshold, full taxation set in⁵.

The mini-jobs reform abolished the hours restriction, expanded the range of exempted earnings up to a threshold of 400 € per month, and introduced a SSC subsidy on monthly earnings between 401 and 800 €. Within this range, SSC start at 4 percent and then increase linearly up to the normal rate of 21 percent at the end of the bracket. Above 800 €, SSC are due on the whole amount of gross earnings. Employers have to pay SSC of 25 percent of the employee's wage for earnings up to 400 €⁶. Between 401 and 800 €, employers pay the normal rate of 21 percent. Earnings up to 400 € are taxed by a flat rate of 2 percent.

Presumably, the mini-job reform was intended to increase employment of persons with low earnings potential. However, eligibility to the SSC subsidy does not depend on a *minimum* number of working hours. Hence, people, like pensioners and students, who want to work only a few hours at a relatively high hourly wage benefit from the subsidy as well as people with low earnings capacity. There is little financial incentive to take up a mini-job for people receiving means-tested benefits because of the extremely high social assistance withdrawal rate under status quo conditions (see section 3.2). Furthermore, for married couples the marginal tax rate may become very high due to the interaction of the SSC subsidy with the tax system (Steiner and Wrohlich 2004). A negative labor supply effect of the mini-jobs reform could arise if already employed people reduced their working hours to take advantage of the SSC subsidy and preferential tax treatment of the new mini-jobs. Overall, it seems rather uncertain whether the mini-jobs reform will contribute to the political goal of reducing unemployment in Germany.

⁵ Actually, there briefly existed another piece of legislation between March 2002 and March 2003, called the “*Mainzer Modell*”, which was also meant to subsidize SSC and was initially intended to be terminated by 2006. Under this scheme, the subsidy also depended on family status and the number of children, and social assistance was not withdrawn for people holding such jobs.

⁶ For people employed by private households, the employers' SSC are only 12 percent of the employee's wage.

This pessimistic view on the potential labor market effects of the mini-jobs reform is also confirmed by the results presented in Steiner and Wrohlich (2004) and summarized in Table 1 below. These results, based on a behavioral microsimulation study, show that the total participation effect is estimated to amount to only about 53 thousand persons, which corresponds to roughly 36 thousand full-time equivalents. The total hours effect, however, is negative, since those people who are already working reduce their working hours; this negative conditional hours effect amounts to 37,500 full-time equivalents. Therefore, in terms of full-time equivalents, the overall labor supply effect of this reform is slightly negative.

Table 1

Estimated Labor Supply Effects of the “Mini-Jobs” Reform (in Thousands)

		Number of persons additionally participating after the reform	Total hours effect (per week)	Hours effect due to additional participation (per week)	Conditional hours effect (per week)
Couples	Women	42	-409	957	-1,367
	Men	9	412	381	31
Singles	Women	2	-82	58	-140
	Men	0	0	0	0
Total		53	-79	1,396	-1,476

Source: Adapted from Steiner and Wrohlich (2004).

These estimates do not include the additional labor supply from students, pensioners and people who may be induced by the reform to take up a second job. Accounting for this type of marginal employment as well, the total labor supply effect of the mini-jobs reform is, therefore, likely to be positive. However, from a labor market perspective a mini-job reform which increases employment of pensioners and students at the cost of reducing total employment of the “main” labor force can hardly be considered a success.

2.4 A Preliminary Assessment

In implementing its Agenda 2010 the German government has introduced some reforms which point into the right direction as their intended chief target is to combat major causes of the problem of long-term unemployment. The introduction of time limits for the entitlement to unemployment benefits by 2006 will reduce long-term unemployment in the medium term. The unification of unemployment assistance and social assistance will reduce unemployment among those who would have been entitled to unemployment assistance under the previous status

quo. This reform is unlikely to help much to strengthen work incentives for recipients of social assistance, however. It seems also unlikely that the new Personal Service Agencies will act as an effective work test. The mini-jobs reform may lead to a slight increase in employment in the low-wage sector, but this will come with substantial deadweight loss and a reduction in total working hours of the main labor force.

3. A Social Welfare Reform That Might Work

In the following, I present a reform proposal which, by avoiding some of the described shortcomings of the Agenda 2010, might lead to a reduction in unemployment and an increase of employment in low-wage labor markets, although at some economic costs. I will first describe the reform proposal in some detail, then discuss the financial incentives induced by the reform proposal, and finally summarize results from a behavioral simulation study on the labor supply, employment and fiscal effects of the proposed reform.

3.1 The Reform Proposal

The proposed welfare reform contains the following three basic components⁷:

- (i) The integration of social and unemployment assistance into one social welfare payment with entitlement conditional on the means test currently applied to social assistance.
- (ii) The level of social assistance for “employable” persons who choose not to work is reduced substantially. All people receiving social assistance are offered a full-time public works job at the current social assistance level.
- (iii) Incentives to take up work shall be improved by a combination of a reduction of the social assistance withdrawal rate and an earnings-related tax credit.

The first component of this welfare reform proposal is probably the least controversial one, at least among economists. Since social assistance and unemployment assistance are both means tested social welfare payments, there is really no reason to differentiate between the two. Furthermore, a major aim of welfare reforms is to reduce long-term unemployment which has been shown to be strongly affected by the availability of unemployment assistance (see Steiner 1997, 2003a).

⁷ This reform proposal, as initially suggested by Steiner (2002), is described in more detail in Steiner and Jacobebbinghaus (2003). Related proposals have been suggested by Sinn et al. (2002), the Scientific Council of the Federal Ministry of Economics (Wissenschaftlicher Beirat 2002), the German Council of Economic Advisors (Sachverständigenrat 2003), and Zimmermann (2003).

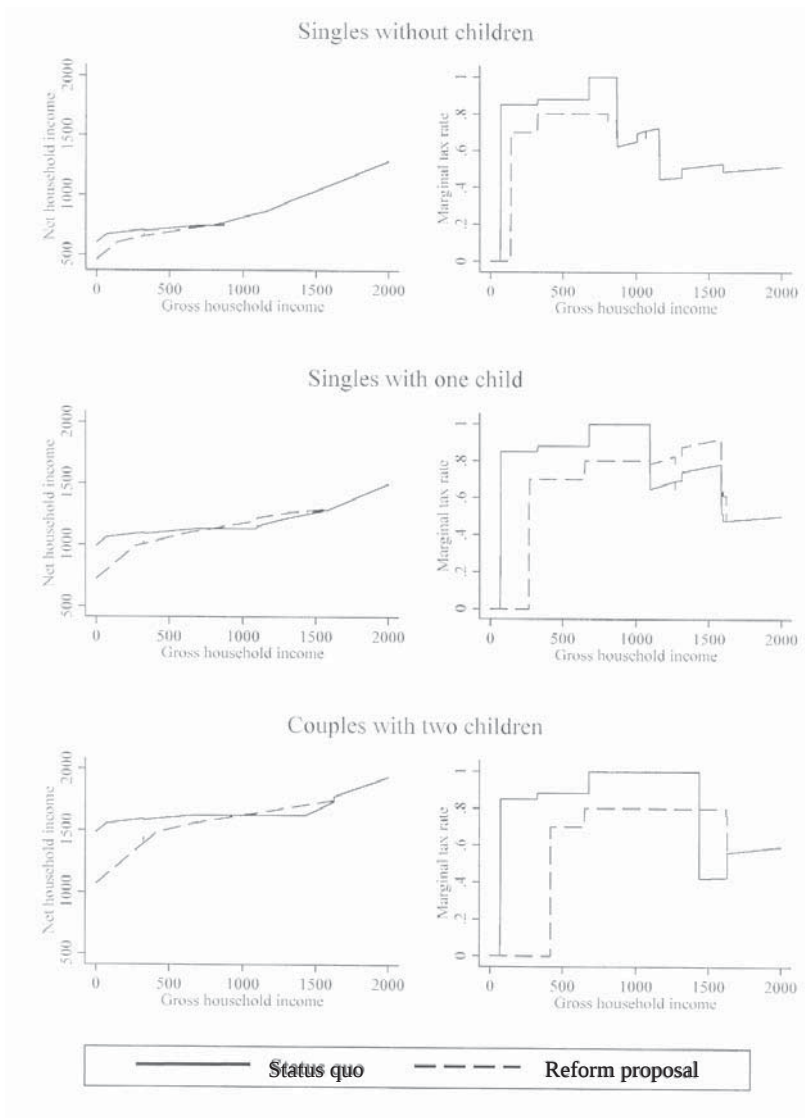
A significant reduction of the social assistance level for “employable” persons who *choose not to work* would also be not very controversial, in principle, since already compatible with existing law. The problem rather is to define what “employable” means, and how a conclusive work test can be implemented. In fact, there is no operational generally accepted definition of “employability”. Here, people will be defined as employable if they are aged between 18 and 65 years, are not severely disabled, are not in full-time education or on maternity leave. From this group, one person per household is excluded if children below the age of 14 years or severely disabled persons living in the household are taken care of. For households with at least one employable person the level of social assistance including housing costs is reduced to about 70% of its previous level if a job offer is rejected.

In order to improve work incentives, the social assistance withdrawal rate is cut to zero until earnings reach the social assistance level under the status quo. In other words, every € earned remains untaxed until the current social assistance level is reached. Individual earnings exceeding that level are taxed at a rate of 70%. Furthermore, similarly to the government’s mini-jobs reform described above, SSC on low earnings are subsidized at a degressive rate; for singles (couples), the tax credit covers a degressive share of the employee’s SSC up to monthly earnings of 820 (1,620) €. In the calculation of social assistance the tax credit is accounted for as earned income. Therefore, people facing the social assistance withdrawal rate of 70% most of the tax credit is compensated for by the reduction of social assistance. The resulting implicit marginal tax rates for various household types are discussed in the next section.

3.2 Improved Work Incentives

The total number of households potentially affected by the analysed welfare reform would amount to almost 2 million, or about 12% of all households in our sample. Counting both heads of households and their partners about 2.7 million or roughly 10% of the respective population would be affected by the welfare reform⁸. About 8% of all households would have less income due to the reform, but for about 4% of all households income would increase. For a large share of households the loss of unemployment assistance under the social reform would be compensated for by social assistance. The average reduction of unemployment insurance in the amount of about 450 € per month therefore translates only into a modest reduction in the level of social assistance of about 30 € in west Germany and about 50 € in the east. Figure 4 illustrates the implications of the reform proposal with respect to net household incomes and marginal tax rates and compares them with the current system for various household types (for the details of the calculations see Steiner and Jacobebbinghaus 2003).

⁸ For a more detailed description of differences between east and west Germany and between various household types see Steiner and Jacobebbinghaus (2003, table 1).



Source: Steiner and Jacobebbinghaus (2003).

Figure 4: Gross Earnings, Net Household Income and Marginal Tax Rates under the Current System and under the Reform Proposal

As shown in the upper part of Figure 4 for *singles without children*, due to the substantial reduction of the social minimum net household income after the reform would always lie below the income level under the status quo until gross earnings reach 870 €, where the two lines coincide. The upper-right panel of Figure 4 shows that up to that earnings level the marginal tax rate under the reform is always below the one for the status quo, and this difference is about 20 percentage points in the range between 680 and 870 €.

The middle part of Figure 4 shows that *single parents with one child* could increase their net household income after the reform by expanding their labor supply: between 785 and 1,620 € gross earnings post-reform net household income would be substantially higher than under the status quo. Marginal tax rates after the reform would be substantially below current levels over a substantial rate, and the maximum marginal tax rate would never reach 100%, in contrast to the current system. However, between 1,100 and 1,620 € marginal tax rates after the reform would be somewhat higher than under the status quo. This is a direct implication of the now markedly lower marginal tax rates in the lower part of the earnings distribution.

In the lower part of Figure 4, the implications of the reform proposal for a *couple with two children* are illustrated. The emerging picture is similar to the one described in the previous paragraph. Net household income increases with rising gross earnings of the household over the range of 1,010 and 1,630 €. Over a wide range, marginal tax rates under the reform are substantially below those prevailing under the status quo. Post-reform marginal tax rates reach the maximum at 80% compared to 100% under the status quo.

To summarize, the described welfare reform would improve incentives to take up work, especially for larger households and single parents. By reducing the social minimum substantially for persons who choose not to work, the marginal tax rate on low earnings can be significantly reduced without shifting the cut-off point for social welfare too far to the right of the earnings distribution. However, the significant reduction of the marginal tax rate at the lower end of the earnings distribution implies that for larger households and single parents the marginal tax rate under the reform is higher than under the status quo over some range of the earnings distribution.

3.3 Labor Supply and Employment Effects

Simulation results for the labor supply and employment effects of the proposed reform are summarized in Table 2 (*c.f.* Steiner and Jacobebbinghaus 2003, table 2). Numbers in the downward flexible wages amounts to about 320 thousand people. As mentioned above, the relatively small difference between these two effects can be explained by the rather low empirical labor supply elasticities prevailing in Germany.

Simulation results in Table 2 show that the employment increase would be mainly concentrated on west Germany. Given the much higher unemployment rate in East Germany, the relatively small labor supply effect there may seem surprising, but can be explained by the relatively small labor supply elasticity of east German wives. In west Germany the labor supply response is distributed fairly evenly between couples and singles, whereas labor supply effects differ by gender within these two groups. For east Germany, differences between the various groups are too small for any meaningful interpretation.

Table 2

Effects of the Proposed Social Welfare Reform on Labor Force Participation (ΔS) and Employment (ΔE) – in 1,000 Persons

	$\Delta S = \Delta E$ (no wage adjustment)	ΔE (with wage adjustment)
West Germany	322	266
Couples	164	137
Men	108	95
Women	56	42
Singles	158	129
Men	53	46
Women	106	84
East Germany	69	57
Couples	36	30
Men	17	14
Women	20	15
Singles	33	27
Men	6	5
Women	28	23
Germany	391	323

Source: Adapted from Steiner and Jacobebbinghaus (2003).

According to the proposed welfare reform, employable people not working in a private sector job can avoid a reduction of social assistance if they take up a public-works job. Steiner and Jacobebbinghaus (2003) try to estimate the expected number of people taking up such a job. The idea is to compare an household's utility evaluated at the working hours the household has to work in order to receive the current social minimum with the same household's utility evaluated at zero working hours and the respective level of the social minimum as stipulated by the reform. Simulations show that in about 1/3 of about 800 thousand households whose social assistance level would be reduced under the welfare reform either the

household head or the spouse would take up a public-works job to avoid the reduction of the social assistance level. Hence, about 2/3 of all affected households would choose not to take up a full-time public-works job and accept the reduction in the social assistance level. This contributes substantially to the positive net balance of the budgetary effects of the welfare reform proposal which amounts to yearly net savings in social expenditures of almost 10 billion € (for the details of the calculations see Steiner and Jacobebbinghaus 2003).

4. Implications for Economic Policy

The starting point of this paper was the recent discussion of social welfare reforms with the aim to increase work incentives in Germany, thereby reducing unemployment of low-skilled workers and increasing employment in the low-wage sector. The dilemma of the German welfare system is the impossibility to retain the high level of the social minimum (for larger households) relative to the average level of earnings, increase incentives to take up low-wage job by reducing the benefit withdrawal rate for social welfare and, at same time, constrain fiscal costs within sustainable limits.

The reason for this fact is that, given a relatively high social minimum and a fairly compressed earnings structure, a social welfare withdrawal rate substantially below 100 percent as it would be required for improving work incentives, implies either significantly higher marginal tax (withdrawal) rates over some part of the earnings distribution or the extension of social welfare eligibility to households with relatively high earnings, or a combination of both. Since the various groups in society are affected quite differently by these reforms, the choice between these unpleasant alternatives depends on the distributional weights assigned to each group. There is no first-best solution to this problem which makes no group worse off.

Given that the reduction of long-term unemployment is the main political goal of labor market and social policy reforms, the German government's Agenda 2010 points into the right direction. However, with the exception of the reduction of entitlement periods for unemployment benefits, the reforms so far implemented by the government under its Agenda 2010 seem unlikely to have a strong impact on unemployment in Germany. The reasons for this pessimistic view are, first, that the social welfare and the mini-job reforms do not set proper work incentives and, second, there is no effective work test reinforcing the potentially positive effects of these reforms.

In this paper, I have presented a welfare reform which might work better than the social reforms of the government's Agenda 2010. The proposed reform has the following components: (i) the integration of unemployment assistance and social assistance; (ii) a substantial reduction of the social assistance level for "employ-

able” persons who choose not to work; (iii) improved incentives to take up work by a combination of a reduction of the social assistance withdrawal rate and an earnings-related tax credit. The empirical results presented in this paper show that the proposed welfare reform would increase employment in the regular labor market by about 300 thousand persons. Another 300 thousand persons entitled to social assistance can be expected to take up a public-works job in order to avoid cuts in the social assistance level.

The introduction of the welfare reform proposal presented here would also lead to a substantial reduction in net social expenditures. The lion’s share of these savings comes from the integration of unemployment assistance into social assistance and the reduction of its level for those not willing to take up a regular or public-works job. Even taking into account the reduction in tax receipts induced by the decline in market wages, the net budgetary effect of the welfare reform will remain substantial. This may even be true if the financial costs for providing public-works jobs for those who are willing to take up but do not find jobs in the private sector are taken into account, depending on how the public-works sector would be organized.

The expected positive effects of the reform proposal come at some economic costs, however. First, the provision of subsidized public-sector jobs for a relatively large number of social assistance recipients, which is likely to be a prerequisite for the political acceptability of the work requirement included in the reform proposal, implies an expansion of the already large public sector. Furthermore, by distorting labor supply decisions of people currently not entitled to social welfare, work requirements may be economically inefficient. Second, the reform proposal also implies that marginal tax rates will increase in some part of the earnings distribution, distorting labor supply decisions of already employed people. Given that there are both losers and winners as well as potential efficiency losses, a decision on the introduction of the proposed social reform has eventually to be based on value judgements.

References

- Akerlof, G. (1978): “The economics of tagging as applied to the optimal income tax, welfare programs, and manpower planning”, *American Economic Review*, 68, 8–19.
- Besley, T. J./Coate, S. T. (1992): “Workfare versus welfare: incentive arguments for work requirements in poverty-alleviation programs”, *American Economic Review*, 82, 249–261.
- Blank, R. (2002): “Evaluating welfare reform in the United States”, *Journal of Economic Literature*, XL, 1105–1166.
- Blundell, R. (2000): “Work incentives and ‘in-work’ benefit reforms: a review“, *Oxford Review of Economic Policy*, 16:1, 27–44.

- Blundell, R. / Duncan, A. / McCrae, J. / Meghir, C. (2000), "The labour market impact of the Working Families' Tax Credit", *Fiscal Studies*, 21(1), 75–104.
- Blundell, R. / MaCurdy, T. (1999): "Labor supply: A review of alternative approaches", in *Handbook of Labor Economics*, edited by O. Ashenfelter and D. Card, 3A, Elsevier, Chapter 27: 1559–1695.
- Buslei, H. / Steiner, V (2003): "Anreizeffekte von Lohnsubventionen: Wie wichtig sind sie für die aktuelle Reformdiskussion", *Vierteljahreshefte zur Wirtschaftsforschung*, 2/2003, 94–108.
- Eissa, N. / Hoynes, H. (1998): "The Earned Income Tax Credit and the labor supply of married couples", Working Paper 6856, National Bureau of Economic Research, Cambridge, Mass.
- European Economic Advisory Group [EEAG] (2002): "Report on the European Economy 2000", CESifo, Munich, Chapter 6: Welfare to Work, 71–86.
- Feist, H. / Schöb, R. (1998): "Workfare in Germany and the problem of vertical fiscal externalities", *Finanzarchiv*, 55, 461–480.
- Franz, W. (2003): "Arbeitsmarktökonomik", 5th edition, Springer.
- Fuchs, V. R. / Krueger, A. B. / Poterba, J. M. (1999): "Economists' views about parameters and policies: survey results in labor and public economics", *Journal of Economic Literature*, 36, 1387–1435.
- Haan, P. (2004): "Discrete choice labor supply: conditional logit vs. random coefficient models", DIW Discussion Paper, 394.
- Hartz-Kommission (2002): "Bericht der Kommission – Moderne Dienstleistungen am Arbeitsmarkt", German Federal Government Publication, Berlin.
- Homburg, S. (2003): "Arbeitslosigkeit und soziale Sicherung", *Vierteljahrshefte zur Wirtschaftsforschung*, 2003/1, 68–82.
- Hotz, V. / Scholz, J. K. (2003), "The Earned Income Tax Credit", in *Means-Tested Transfer Programs in the United States*, edited by R. Moffitt, University of Chicago Press, Chapter 3.
- Meyer, B. (2002): "Labor supply at the extensive and intensive margins: The EITC, welfare, and hours worked", *American Economic Review*, 92(2), 373–379.
- Moffitt, R. (2002): "Welfare programs and labor supply", NBER Working Paper 9168, Cambridge, Mass.
- (2003): "The negative income tax and the evolution of U.S. welfare policy", *Journal of Economic Perspectives*, 17, 119–140.
- Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2002): "Zwanzig Punkte für Beschäftigung und Wachstum", *Jahresgutachten 2002/03*, Wiesbaden.
- Saez, E. (2002): "Optimal income transfer programs: intensive versus extensive labor supply responses", *Quarterly Journal of Economics*, 117, 1039–1073.
- Sinn, H.-W. et al. (2002): "Aktivierende Sozialhilfe. Ein Weg zu mehr Wachstum und Beschäftigung", *ifo Schnelldienst*, 55, 2002/9, 1–52

- Steiner, V. (1997): "Extended benefit – entitlement periods and the duration of unemployment in west Germany", ZEW Discussion Paper, 14, Mannheim.
- (2002): "Kombilohnmodelle – Ein Weg zur Erhöhung der Beschäftigung in Deutschland?", ifo-Schnelldienst, 4/ 2002, 5 – 9.
 - (2003): "Senkung der Arbeitslosenunterstützung: Weniger Arbeitslosigkeit, mehr Effizienz", DIW Wochenbericht, 25, 401 – 408.
 - (2004): "Labour market reforms for more growth", in Issue No. 11 of the series: "More Growth for Germany", edited by Deutsche Bank Research, Frankfurt am Main.
- Steiner, V. / Jacobebbinghaus, P. (2003): "Reforming social welfare as we know it? A micro-simulation study for Germany", Mimeo DIW Berlin, www.fu-berlin.de/wifo/publications/social.pdf.
- Steiner, V. / Wrohlich, K. (2004): "Work incentives and labor supply effects of the 'Mini-Jobs Reform' in Germany", Mimeo, DIW Berlin.
- Wissenschaftlicher Beirat beim Bundesministerium für Wirtschaft und Technologie (2002): "Reform des Sozialstaats für mehr Beschäftigung im Bereich gering qualifizierter Arbeit", Dokumentation des BMWi, 512, Berlin.
- Zimmermann, K. F. (2003): "Beschäftigungspotenziale im Niedriglohnsektor", Vierteljahrshefte zur Wirtschaftsforschung, 1, 11 – 24.

Social Welfare Reform and the Low-Wage Labor Market in Germany: What Works and What Doesn't?

Comment

By Norbert Berthold*

From welfare to workfare – more decentralized solutions are needed.

Germany's low-qualified jobless are in a predicament. Globalization, structural change and technological progress lead to unemployment for more and more low-qualified people, also due to the fact that too much is lumped together in regard to wage policy and collective bargaining policy. Social security services granted in case of unemployment are pulling in the same direction: they complicate the re-entry into a regular employer-employee relationship. Social assistance plays an important part. By offering close to a minimum wage, it solidifies the lower limit of the structure of qualificatory wages and therefore prevents an urgently needed low-income sector from emerging. Furthermore, due to the fact that the necessary gap between remuneration and social assistance is virtually nonexistent for many households and the fact that tax allowance transfers are reduced by 85 or 100% after a certain period, incentives for low-qualified transfer beneficiaries to (gradually) re-start a regular employment are severely impaired. Many people on welfare are therefore caught in a trap. (Boss 2002, Schneider/Lang/Rosenfeld et al. 2002).

In his contribution, Viktor Steiner rightly emphasizes that the incentivizing "stick" alone, as defined by a reduction of payments and strict sanctions on lack of willingness to co-operate and on offences by transfer recipients, does little to achieve the desired ends. Only in combination with actual work options and corresponding positive incentives are the "carrots" – a national increase in employment and a reduction in dependence on transfers – achievable. Thus, what must be done is to stimulate work supply and demand for low-qualified employees. The so-called Hartz Laws on Agenda 2010 have produced no great successes in this regard. It is true that the fusion of unemployment assistance and social assistance into the so-called "unemployment benefit II" is long overdue and makes good economic sense. However, at least two points spoil its effect. First, the generosity of transfer services for persons fit for work is not appropriate to the economic

* Universität Würzburg, E-mail: norbert.berthold@mail.uni-wuerzburg.de.

reality, nor is the rate of reduction of social security transfers reduced decisively. Consequently the reorganization of the so-called “Mini-jobs” and “Midi-Jobs” also clutches at thin air and barely reaches the true target group (Berthold / von Berchem 2003a). Just a temporally limited contribution for employees if they take up work supports employment – if the respective person in charge considers this to make sense. Second, the opportunity has not been seized to use the power of decentralized solutions or to provide local authorities with decisive parameters.

Viktor Steiner’s contribution aims at the first point and puts forward a proposal for sustainable improvement in incentives to work and a reduction of unemployment resulting from minimum wages brought about by social security services. Substantial lessening of transfer payments to people unrestrictedly able to work as well as a decisive reduction in the marginal burdening of additional earned income by a combination of short reduction rates of social security transfers and tax credits independent of income would beyond doubt lead to a better employment environment for low-qualified employees from the perspectives of both labor supply and demand. The specified possible effect on employment basing on empirically estimated elasticities of labor supply and labor demand, as well as the budgetary impact, seem to be plausible. It is important to note that state-aided employment opportunities that should be offered to each welfare recipient who is able to work also entail problems. This is because these jobs are not only costly in maintenance, but also raise the basic dilemma that on the one hand, they should not crowd out regular employment and on the other, a certain marketability is necessary in order to avoid additional reductions in job prospects for the regular participants in the program (Berthold / von Berchem 2003b).

The second point – the question of legal and institutional competence – is not touched on by Viktor Steiner. As it is an important topic for a comprehensive reform proposal, I will cover it here briefly. The consequences of long-term unemployment and poverty are brought to bear most immediately – apart from the families of those affected – at the communal level. Thus, the communities are interested in avoiding these consequences. Experience in Germany and abroad shows that in this context, decentralized solutions perform well, and are usually superior to centralized approaches. The dominant strategy is decentralization of the responsibility for long-term unemployed people and their re-integration at the communal level (Berthold / Fehn / von Berchem 2001, Klös / Peter 2001, Berthold / von Berchem 2002a). Such decentralization must be done comprehensively, however, and it must not be restricted to implementation and control of regulation and financing.

The compromise proposed by the mediation committee and now implemented, containing the idea that in case of financial compensation communities can be used as an extended arm of the federal law to fight against long-term unemployment and poverty, is no alternative to a genuine decentralization of the response. It does not fulfill the demands of an efficient labor market and social policy. From a

public choice point of view, it is hardly surprising that – in negotiations dominated by the federal government where the communities are represented by the Länder, i.e. practically left aside – very little central power is given up and communities even have to fear being cheated financially. Looking at this economically, the compromise of optional “Organleihe” is hardly target-orientated. For an extensive use of the opportunities to solve the problems, it is urgently necessary that more authority be decentralized, including responsibility for implementation and much more decision-making and regulation. Responding to the complexity of unemployment and poverty is possible only by means of decentralized experiments and tailor-made solutions on the spot. The same applies to the use of institutional competition. Severe and binding constraints in the form of central regulations without basic possibilities for a break prevent this from the very outset.

The modification in the reduction of transfers linked to taking a paid job and a general reduction in transfer payments to people who are unrestrictedly fit to work are necessary steps for a promising reform, but still far from sufficient in themselves. The unconditional willingness to work must be required as a condition for public transfers, and the efforts of individuals must be supported as well as possible. The balance of sticks and carrots, of demanding and encouraging, cannot be defined completely in advance. On the other hand, however, centrally packing a binding bag of measures and its exhaustive conversion is neither useful nor necessary. It would be much better to start granting much more room to maneuver in the organization of social transfers to communities since they are the territorial authorities that are originally in charge of social assistance. Because of the existing gap between social transfers and incomes that can be reached on the spot, and the fact that the transfers should both be geared to local cost of living and furthermore take into account that prospects of getting a new job for unemployed differ regionally, the extent of the transfer should also be part of the communities’ sphere of authority. Apart from this, communities should be allowed to determine the exact organization of the reduction of transfers linked to taking a paid job in the period of time. The same goes for subsidies of wages. Finally it should be the business of the communities what kind of demands should be made of people on social security, and what kind of sanctions should be imposed in case of uncooperative behavior.

The most successful strategies in such a system will be worked out through a process of experimentation at the decentralized level. These programs will better conform to local circumstances and it will be easier for them to achieve custom-made solutions to individual problem cases. However, if the communal pool of opportunities for solving problems is used in the desired way, the decentralized competences for organization require at least two things: First, it is compellingly necessary that, apart from the authority for regulation and enforcement, the authority for financing this policy sphere also be handed over. Central financing of locally provided payment nips any kind of competition in the bud. Only if the communities are given a sufficient share in the financing of their own programs can it

be ensured that there will be a quest for the best methods, and that the consequences of individual measures will be completely internalized. Innovated experimentation in the fight against unemployment and poverty therefore requires more flexibility and autonomy, also regarding the communal receipts. Second, it must no longer be possible for claims to be made on unemployment insurance by means of artificially created, temporary communal jobs. “Switchyards” that are a result of this practice should shut down at once – for the good of the unemployed. An intelligent fusion of unemployment assistance and social assistance, in combination with a genuine reform of communal finance, could accomplish the necessary framework (a comprehensive proposal for reform in this vein can be seen at Berthold / von Berchem 2002b, 2003b).

References

- Berthold, N./Fehn, R./von Berchem, S. (2001): “Innovative Beschäftigungspolitik – Wege aus der Strukturkrise”, Bad Homburg.
- Berthold, N./von Berchem, S. (2002a): “Sozialhilfe im wettbewerblichen Föderalismus: Erfahrungen der USA, Lehren für Deutschland”, in Bürgernah fördern und fordern – Konzepte für eine effiziente Arbeitsvermittlung und Grundsicherung, edited by G. Raddatz, Berlin, 91 – 114.
- (2002b): “Kampf gegen Arbeitslosigkeit und Armut: Markt, Staat und Föderalismus”, Berlin.
- (2003a): “Job-AQTIV, Hartz, Agenda 2010 – Aufbruch zu neuen Ufern oder viel Lärm um Nichts?”, in Damit Hartz wirkt!, edited by G. Raddatz and G. Schick, Berlin, 31 – 52.
- (2003b): “Die Sozialhilfe zwischen Effizienz und Gerechtigkeit – wie kann der Spagat gelingen?”, in Theorie der sozialen Ordnungspolitik, edited by: N. Berthold and E. Gundel, Stuttgart, 137 – 157.
- Boss, A. (2002): Sozialhilfe, Lohnabstand und Leistungsanreize, Berlin.
- Klös, H.-P./Peter, W. (2001): “Strategien einer aktivierenden Sozialpolitik”, in Fördern und Fordern: ordnungspolitische Bausteine für mehr Beschäftigung, Institut der deutschen Wirtschaft, Cologne, 177 – 210.
- Schneider, H./Lang, C./Rosenfeld, M. et al. (2002): “Anreizwirkungen der Sozialhilfe auf das Arbeitsangebot im Niedriglohnbereich”, Halle (Saale).

A Conceptual Framework for the Evaluation of Comprehensive Labor Market Policy Reforms in Germany

By Michael Fertig* and Jochen Kluge**

Abstract

Over the last year the German government has introduced a comprehensive set of labor market policy reforms, the so-called *Hartz reforms*, which aim at a significant reduction of unemployment. To this end, (a) many of the existing instruments of active labor market policy are modified considerably, (b) a set of new instruments is introduced, and (c) the administrative framework in which these measures operate is changed substantially. In order to be able to judge the success of these measures by the end of the current legislative period in mid-2006, the government has asked academic experts to set up an evaluation concept capable of generating reliable empirical evidence by that date. The task is therefore to develop a ready-to-implement concept for the evaluation of the full set of reforms in their entirety, as well as each instrument on its own, facing substantive constraints regarding data availability and a short time horizon. This paper presents such a concept. We discuss essential guidelines for an ideal evaluation design, conceptual and practical difficulties that arise in the context of evaluating the Hartz reforms, and ways to overcome these obstacles. After detailing the three main analytical steps – analyses of effectiveness, efficiency, and implementation and process analysis – we present the concrete evaluation design, specific methods applicable to particular instruments, and a sampling scheme for collecting the required data. In addition to the fact that our concept is directly implementable, it also has the advantage of being extensible for future labor market policy evaluations.

JEL Classification: J0

Keywords: Program Evaluation, Identification, Active Labor Market Policy, Policy Reform, Evaluation design.

* RWI-Essen and IZA-Bonn

** RWI-Essen and IZA-Bonn. *Corresponding author:* Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI-Essen), Hohenzollernstr. 1–3, 45128 Essen, Germany, Fax: +49–201–8149–236, E-mail: kluge@rwi-essen.de.

The authors are grateful to Helmut Apel, Werner Friedrich, Helmut Hägele, Christoph M. Schmidt, Ronnie Schöb, participants at the BMWA workshop “A Concept for the Evaluation of the Suggestions of the Hartz-Commission”, December 2003, Berlin, and participants at the 2004 ARGE meeting in Berlin for helpful comments.

1. Introduction

Every year Germany spends several billion euros on active measures of employment promotion, with the explicit intention to contribute to the reduction of unemployment. Yet, unemployment has turned out to be a persistent problem throughout the last two decades, raising doubts as to the effectiveness of such employment promotion measures. In line with a general tendency to redefine and modernize various kinds of administrative institutions, several major reforms of employment promotion policy have been launched since 1998. The latest are the so-called *Hartz reforms*: These were first introduced at the end of 2002, and the legislation process continued through to the end of 2003.

The Hartz reforms constitute a comprehensive reform of labor market policy. Indeed, it seems safe to argue that this is the most far-reaching reform endeavor in Germany in recent decades, and the government pursues an ambitious set of objectives with these reforms. They are supposed, in particular, to contribute to a significant reduction of unemployment by a quicker and more sustainable job placement, as well as to the creation of new jobs and employment opportunities. To this end, (a) many of the already existing instruments of active labor market policy in Germany are modified considerably, (b) a set of new instruments is introduced, and (c) the framework in which these measures operate is changed substantially. The latter does not only apply to the new administration and management of labor market policy within the *Federal Employment Agency* (“Bundesagentur für Arbeit, BA”), but also, for instance, to tighter sanctioning measures that aim at imposing more pressure on unemployed workers to take up jobs.

The success of these reforms, however, is anything but guaranteed *a priori*. To learn about policy success, i.e. policy effectiveness, clearly requires systematic evaluation of any implemented measure. This task is especially challenging if policy changes are multifaceted, detailed, and comprehensive, as is the case for the Hartz reforms.

Over the last decades, European policy makers, and German policy makers alike, have not shown much interest in initiating or considering systematic evaluation of policy measures – unlike the US, where scientifically sound evaluations have commonly accompanied policy interventions at least since the 1960s (cf. Kluve and Schmidt 2002 for details on the US-European comparison in this regard). Fortunately, and only very recently, also a European “evaluation culture” has started to emerge, and Germany is no exception. In the particular case of the Hartz reforms, the German parliament (“Bundestag”) explicitly connected the evaluation task to the implementation of the reforms – with the objective to be informed, on a scientifically solid basis, about reform effectiveness before the end of the current legislative period in mid-2006. This decision led to the novel situation that, for the first time, academic evaluation experts are involved from the very beginning in the implementation and evaluation of policy reforms in Germany, which in turn will result in the first systematic evaluation study for such measures.

As the Hartz reforms are currently being set into practice, the task with which the government approached academics can be summarized as follows: to develop a ready-to-implement concept for the evaluation of the full set of reforms in their entirety, as well as each instrument on its own. The main restriction regarding this task is that the concept needs to allow for generating of robust evidence until mid-2005 (1st report) and mid-2006 (Final report), respectively.

This paper presents the main features of the concept that fulfills this task, i.e. we suggest a conceptual framework for a comprehensive evaluation of the effectiveness and efficiency of the Hartz reforms.¹ Specifically, we discuss the *conceptual* and *practical* problems of evaluating current labor market reforms in Germany. We will devote particular attention to the following research questions on the individual, i.e. micro, as well as the macro level:

1. How can we assess the *effectiveness* of labor market reforms?
2. How can we assess the *efficiency* of labor market reforms?
3. How can we determine the *reasons* for the estimated effectiveness and efficiency of the reforms?

In this endeavor, it will be of prominent relevance to take into account the intimate connection between changes in the *general framework* of labor market policy and the *modification and introduction of particular instruments*. The interaction between these two reform components induces the difficulty to *isolate* the impact of a specific policy change. The latter, however, is a prerequisite to identify the *causal effect* of any intervention. Furthermore, many of the reforms induce a *universal treatment* situation, i.e. every unemployed individual is affected by these changes, rendering the construction of a valid comparison group difficult. We will discuss these difficulties in detail and provide suggestions for ways to overcome these obstacles.

The paper is organized as follows. Section 2 establishes the context of evaluation. It discusses objectives and elements of the Hartz reforms, presents guidelines for a suitable evaluation concept, and details the conceptual and practical problems that arise for the evaluation in this specific context. Section 3 focuses on the general methodological framework for the evaluation, i.e. how program effectiveness and program efficiency can be identified. In Section 4 we discuss the concrete evaluation design and delineate specific methods applicable to the elements of the Hartz reforms. This includes details on the appropriate sampling procedure. Section 5 concludes.

¹ The paper is based on a pilot study by RWI-Essen and ISG Köln (cf. Fertig et al. 2004) that develops the evaluation concept for the Hartz reforms *in extenso*. The pilot study was prepared on behalf of the German Federal Ministry for Economic Affairs and Labor ("Bundesministerium für Wirtschaft und Arbeit, BMWA"). A second pilot study by ZEW Mannheim also intends to formulate such a concept.

2. The Hartz Reforms: Establishing the Context of Evaluation

2.1 Objectives of the Hartz Reforms

The new measures and modifications of labor market policy in the framework of the Hartz reforms aim at quick and sustainable job placement, the creation of new employment opportunities, and the simplification of moving into employment. This serves the overall objective of reducing unemployment in Germany. Hence, success or failure of the Hartz reforms must be judged in relation to the claim that the reforms contribute in a sustainable way to combating unemployment. The evaluation concept must therefore focus on the employment performance of unemployed individuals as the central outcome of interest.

The universe of labor market policy reforms within the Hartz framework can be divided into two major groups.

I. Modification and Implementation, respectively, of particular – i.e. aimed at unemployed individuals or individual enterprises – instruments of labor market policy.

- Direct Wage Subsidies (“Eingliederungszuschüsse”, EGZ)
- Self-employment start-up subsidies (“Ich-AG”, “Überbrückungsgeld, ÜG”)
- Qualification measures (“Förderung der beruflichen Weiterbildung”, FbW)
- Regulations regarding low-wage jobs (“Mini-Jobs”)
- Regulations regarding low-wage jobs subject to social insurance contribution (“Midi-Jobs”)
- Personnel Service Agencies (“Personal-Service-Agenturen”, PSA)
- Integration measures provided by a third party (“Integrationsverträge”)
- Reform of Social Plan Measures and Structural Short-Term Allowance (“Umgestaltung präventiver Instrumente, Transferleistungen”)
- Reform of regulations regarding Temporary Help Service Workers (“Arbeitnehmerüberlassungsgesetz”)
- Reform of regulations regarding temporary contracts with old workers and wage subsidies for old workers
- Placement vouchers for private agencies (“Vermittlungsgutscheine”)

II. Modifications in the general framework of labor market policy.

(a) Modifications regarding objectives, administration, and steering system.

- Merging of measures of direct job provision in the public sector (“Arbeitsbeschaffungsmaßnahmen”, ABM, “Strukturanpassungsmaßnahmen”, SAM)
- Simplification of calculation of unemployment benefits

- Output-oriented steering system (“Ergebnis-orientiertes Steuerungskonzept”)
- “Job-Center” as uniform contact point for all job seekers

(b) *Modifications regarding benefit regulations.*

- Reducing of duration of unemployment benefit entitlement
- Tightening of benefit sanctions (“Sperrzeitregelung”, “Zumutbarkeitsvorschriften”)
- Merging of unemployment assistance and social assistance (“ALG II”)
- Uniform regulation regarding employment duration necessary for benefit entitlement
- Social security coverage of persons in compulsory military and community service
- Obligation for early unemployment registration

Table 1 presents a detailed overview of these measures in terms of their respective core features and objectives. Clearly, it is a challenging task to evaluate individual policy measures separate from each other given interactions and overlap in their contents and objectives. The evaluation aims at two things: First, an assessment of the overall effect of the Hartz reforms in their entirety. Second, to isolate and quantify the effect of particular elements of the reforms.

2.2 Guidelines for a Suitable Evaluation Concept

As previously outlined, the task is to develop a ready-to-implement framework for the evaluation of the full set of reforms in their entirety as well as each instrument on its own. This task faces the restriction that the evaluation design needs to allow for generating of robust evidence until mid-2005 and mid-2006, respectively. Hence, we believe that the concept appropriate to fulfill this task must follow a set of guidelines: Practicability, focus, consistency, multidisciplinary, robustness, and extensibility. Specifically:

1. *Practicability*: Given the constraints of a short time horizon and limited data quality and availability, the evaluation concept must be *practicable*. The actual implementation must allow for generating of robust evidence until mid-2005 and mid-2006.

2. *Focus*: Given the multitude of objectives and sub-objectives of labor market policy, as well as the multitude of specific policy changes constituting the Hartz reforms, the evaluation concept must be *focused*. Since reduction of unemployment is the main goal of both the Hartz reforms and labor market policy in general, the evaluation concept must necessarily focus on the effect of the Hartz reforms on the employment situation of unemployed individuals. For some policy measures, the secondary objective “employability” rather than direct integration into

Table 1: Elements and Objectives of Hartz Reforms

Measure	Key Elements	Objectives
<i>I. Introduction of New Instruments, and Reform of Existing Instruments of Active Labor Market Policy</i>		
Direct Wage Subsidies ("Eingliederungszuschüsse", EGZ)	Reduction of the number of wage subsidies to two types: (i) EGZ for unemployed with integration problems and (ii) EGZ for unemployed with disabilities; reduction and unification of benefit duration and amount	(i) Integration of unemployed by temporarily subsidizing their wage; (ii) simplifying the implementation of the instrument and increasing its attractiveness; (iii) increasing the efficiency of the instrument
Self-Employment Start-Up Subsidies ("Ich-AG" and "Überbrückungsgeld", ÜG)	Ich-AG: Temporary (max. 3 years) subsidy for newly self-employed if expected yearly income does not exceed 25,000€; amount of subsidy declines every year; application must be renewed each year ÜG: Subsidy for newly self-employed for six months amounting to unemployment benefits or assistance	(i) Integration of unemployed by supporting self-employment; (ii) creation of additional jobs by newly established companies
Qualification Measures ("Förderung der beruflichen Weiterbildung", FbW)	Introduction of vouchers ("Bildungsgutschein") for choice of provider of qualification measure; certification of providers by external agencies	(i) Integration of unemployed by improving their qualification; (ii) more competition among providers of qualification measures; (iii) higher quality of measures; (iv) reduction of administrative effort
Integration Measures Provided by Third Party ("Integrationsverträge")	Possibility to award integration measures to providers by call for tenders; honorarium for especially successful measures; output-oriented steering of measures	(i) Integration of unemployed by customized and innovative measures; (ii) increasing competition among providers of integration measures
Personnel Service Agencies ("Personal-Service-Agenturen", PSA)	Implementation of at least one PSA in every local labor office district; PSA employs unemployed workers with integration problems for a maximum period of 12 months and engages as a temporary help service company; during periods in which the former unemployed does not work, the PSA is supposed to provide adequate qualification measures	(i) Integration of unemployed by temporary help service jobs; (ii) increasing the acceptance and quality of temporary help service work

Table 1 (continued)

Measure	Key Elements	Objectives
<i>1. Introduction of New and Reform of Existing Instruments of Active Labor Market Policy – cont'd</i>		
Special Wage Subsidies for Old Workers (“Entgeltssicherung für ältere Arbeitnehmer”) and Reform of Regulation Regarding Temporary Contracts With Old Workers	Special Wage Subsidy: Temporary allowance for workers aged 50+ if they take up employment with a lower wage compared to their last job before unemployment; eligibility depends on a residual claim for unemployment benefits of at least 180 days; application by worker required Temporary Contracts: Reduction of age limit for temporary work contracts from 58 to 52	(i) Provision of incentive to take up jobs with a lower wage for old workers by partial compensation of loss in income (and pension claims); (ii) integration of old workers by facilitating temporary employment
Reform of Regulations Regarding Temporary Help Service Workers (“Änderungen im Arbeitnehmerüberlassungsgesetz”)	Prohibition of synchronization, repeated dismissal and recruitment, and maximum duration of temporary help service jobs abolished; easing of exemptions for construction sector	(i) Integration of unemployed by temporary help service jobs; (ii) increasing the acceptance and quality of temporary help service work (see also PSA)
Transfer Measures and Transfer Short-Time Allowance (replacing former social plan measures and structural short-time allowance, “Umgestaltung präventiver Instrumente in Tranferleistungen”)	Transfer measures (former social plan measures): All measures aiming at the integration of workers in danger of becoming unemployed because their company will be closed; employer has to take over 50% of costs; transfer measures must be implemented by external provider who is obliged to have a quality assurance system; parallel promotion by other instruments no longer possible Transfer Short-Time Allowance (former structural short-time allowance): Profiling measure necessary for eligibility; reduction of maximum duration of entitlement to 12 months (instead of 24); employer is committed to provide placement assistance or adequate qualification measures	(i) Facilitate transition from work to work by a better coordination of instruments; (ii) avoidance of transitory unemployment periods; (iii) abolish former practice of early retirement via structural short-time allowance
Placement Vouchers for Private Agencies (“Vermittlungsgutscheine”)	Workers unemployed for three months or more can apply for a placement vouchers for a private agency; this voucher is valid for three months; private agencies receive honorarium depending on integration success	(i) Integration of unemployed by utilizing private placement services; (ii) increasing competition for labor offices

Table 1 (continued)

Measure	Key Elements	Objectives
Mini- and Midi-Jobs (“Ger- ingfügige Beschäftigungsver- hältnisse und Einführung ei- ner Gleitzone in der Sozial- versicherung”)	Mini-Jobs: Jobs with wages up to 400€ per month are exempted from income tax and social security contribu- tions for the employee; employer has to pay a 25% flat- rate (taxes and social security contributions); this regula- tion also applies for secondary jobs; voluntary additional contributions to pension system possible; Mini-Jobs for private households receive privileged status; unemployed workers holding a Mini-Job can earn up to approx. 200€ without losing part of their benefit payments Midi-Jobs: Partial reduction of social security contribu- tions for employees for jobs between 401 – 800€; social security contributions increase linearly from 4.25% to the full rate of contribution (approx. 21%)	(i) Integration of unemployed by providing incentives to take up a Mini- or Midi-Job which is supposed to serve as a steppingstone into regular employment; (ii) increasing the flexibility of employment in low-paid jobs; (iii) reduc- tion of non-wage labor costs; (iv) subsidizing low-income sector; (v) reduction of illegal and shadow work (esp. in private households); (vi) reducing bureaucratic effort for employers; (vii) creation of additional opportunities to earn money for workers in low-paid jobs
II. Modifications Regarding Objectives, Administration, and Steering System		
Merging of Measures of Di- rect Job Provision in the Pub- lic Sector (“Zusammenle- gung von ABM und SAM”)	Merging of ABM and SAM into ABM-new; reduction of administrative complexity; participants are no longer cov- ered by social security and therefore participation does not generate new benefit eligibility; objective of ABM- new no longer integration of unemployed, but increasing employability	(i) Increasing employability of unemployed; (ii) avoid- ance of crowding-out effects and benefit churning
Output-oriented Steering System (“Ergebnis-ori- entiertes Steuerungskonzept”)	Implementation of target agreements and contract man- agement system for the steering of labor market policy between the different levels within the BA	(i) Faster and more sustainable integration of unemployed by a more flexible and individually targeted labor market policy; (ii) conversion of labor offices into customer-or- iented service centers
Job-Center (“Job-Center als einheitliche Anlaufstelle für Arbeitssuchende”)	Implementation of a joint center for all recipients of ALG II (see below) providing them with all kind of services necessary for labor market integration	(i) Activation of unemployed; (ii) counseling and moni- toring of job seekers; (iii) co-ordination of placement and advisory activities

Table 1 (continued)

Measure	Key Elements	Objectives
<i>III. Modifications Regarding Benefit Regulations</i>		
Reduction of Duration of Unemployment Benefit Entitlement	Reduction of duration to 12 months (for workers aged 55+: 18 months); transitory period until February 2006	(i) Reduction of unemployment duration; (ii) avoidance of practice of early retirement
Tightening of Benefit Sanctions ("Sperrzeitregelung", "Zumutbarkeitsvorschriften")	Differentiated and unified system of temporary refusal to pay benefits in response to misconduct by unemployed; reversal of the burden of proof from labor office to unemployed Unemployed without family are obliged to move to another region if they receive a job offer	(i) More flexibility in sanctions system; (ii) provision of incentives to take up jobs by reducing benefits; (iii) increasing geographical mobility of unemployed
Merging of Unemployment Assistance and Social Assistance (ALG II)	ALG II as new benefit system for former unemployment assistance or social assistance recipients on the level of former social assistance; stepwise transition from unemployment benefit (ALG I) to ALG II by temporary extra-allowance; possibility to promote all recipients of ALG II by measures of ALMP; transitory period of 6 months for former recipients of unemployment assistance	(i) Integration of long-term unemployed by providing incentives to take up employment (lower level of benefits for former recipients of unemployment assistance and lower earnings threshold for former recipients of social assistance); (ii) reduction of expenditures for long-term unemployed
Obligation for Early Unemployment Registration	Obligation to register as unemployed immediately after notice of dismissal or 3 months prior to the end of a temporary work contract; upon violation: reduction of benefits possible	(i) Speeding-up of placement process; (ii) avoidance of begin of unemployment spell

the labor market has been formulated by policy makers as the outcome measure of interest – clearly, this outcome would be somewhat more difficult to measure and assess in the evaluation.

3. *Consistency*: The evaluation concept must offer a *consistent* framework for the analysis of well posed evaluation questions; a framework that allows for comprehensive answering of these evaluation questions and for generating of robust and convincing evidence. At the same time, given the heterogeneity of policy measures within Hartz, this consistent framework must be sufficiently flexible to accommodate tailor-made evaluations of specific policies. The concept must be based on recurrent steps of analysis for each policy measure, following three central questions:

- a) What are the effects of Hartz interventions, i.e. are these policy measures and policy changes effective?
- b) What costs have been generated by these interventions relative to the effects, i.e. are these policy measures efficient?
- c) What reasons can be identified for the observed effects?

4. *Multidisciplinary*: For a comprehensive assessment of reforms within a consistent analytical framework, the evaluation concept must follow a *multidisciplinary* approach. This requires a balanced combination of sound theoretical analysis, econometric competency, and extensive knowledge of implementation and process analysis.

5. *Robustness*: The evaluation concept must encompass an analytical approach that – given time and data constraints – allows for generating of *robust* evidence. This requires an intimate connection of available data and empirical method. Application of highly complex methods that do not fit the data, and that may generate results that are highly sensitive to specification changes, is certainly counterproductive.

6. *Extensibility*: The evaluation concept must allow for an *extension* of the evaluation horizon beyond 2006. Given the late coming into effect of some policy measures within Hartz, and the important question regarding long-term effectiveness, a continuation of the evaluation beyond 2006 is both sensible and necessary. Hence, the evaluation concept must be transferable in a straightforward manner to later evaluation dates.

In the light of these background guidelines, the main elements of the appropriate concept for evaluating the comprehensive Hartz reforms result in the following procedure in practice.

- Compiling and reporting of main statistical data (Monitoring function).
- Analysis of the improvements of services provided by the employment administration (employment offices) as the main instrument for increasing the effectiveness of labor market interventions.

- Microeconometric analysis of the instruments of Active Labor Market Policy that are implemented or modified within the Hartz framework. This analysis focuses on establishing the effectiveness and efficiency of policy measures regarding the employment performance of unemployed individuals.
- Macroeconomic analysis of labor market policy reforms, taking into account regional, cyclical, and other factors.

A major objective of the Hartz reforms concerns the provision of “modern services” by the employment administration. The fundamental hypothesis is that the effectiveness of a particular instrument of labor market policy does not only result from the instrument itself, but also from the counseling, informing, program allocation, and monitoring provided by the employment offices. Quite clearly, the effectiveness of a training program depends on the extent to which program contents are congruent with market demands, on the type of quality assurance used by employment offices, and on the type of participants selected for a specific program. Therefore the evaluation concept encompasses as central elements the consideration of the quality of service provisions, i.e. essentially the quality of labor offices, and the implementation of the new control system within the employment offices.

2.3 Conceptual and Practical Problems of the Evaluation

Any evaluation study that aims at identifying the causal effect of a policy measure must isolate the intervention to the best possible extent, in order to avoid measuring the effect of various interventions overlaying each other. If there are indeed various measures implemented at the same time – as is obviously the case of the Hartz reforms – then it is the task of the evaluation design to guarantee the identification of the isolated effect. This requires the construction of a genuinely comparable comparison scenario. The comparison scenario aims at answering the counterfactual question, what would have happened (in terms of the outcome of interest) if the policy measure had not been implemented? The difference in the outcome between the factual scenario – the policy measure being implemented – and the counterfactual scenario then measures the causal effect of the intervention. Quite intuitively the construction of the counterfactual scenario has to follow the principle “compare the comparable”, since ideally the factual and counterfactual states should differ only in the policy intervention being implemented or not (and associated outcomes), and nothing else.

The evaluation of the Hartz reforms therefore faces *two conceptual problems*.

I. The intertwining of changes in the general framework of labor market policy and individual instruments of labor market policy

As discussed above, some measures of the Hartz reforms regard the general objectives, administration and steering system of labor market policy. Specifically,

this concerns the new outcome-oriented steering system, which aims at simplification of administrative procedures as a means to let employment office staff focus more intensely on their main task, job placement of unemployed individuals. If this new system were indeed to work, and, for instance, employment office staff could allocate more time to support unemployed workers and select measures of active labor market policy suitable for them, then this would clearly also affect the effectiveness of such programs. Hence, the modification in the general framework overlays any comparison of the reintegration performance of an active labor market program before and after the reforms. This methodological problem would be even more severe, if actual support intensities under the new framework were to vary across employment offices.

It is possible, and necessary, to appropriately reflect the role of the concrete implementation of the new steering system in the employment offices on the reintegration performance of individual programs: Each evaluation study must include as control variables indicators for the quality of service provision for individual employment offices, in order to avoid potentially severe bias in the impact estimates (“omitted variable bias”). Such indicators for the quality of labor offices comprise, for each employment office, the number of job search assistants relative to unemployed individuals, the amount of online job search facilities relative to unemployed individuals, the average duration of counseling interviews, etc.

II. Several specific policy interventions constitute “universal treatments”

Some policy measures – such as the Mini- and Midi-Jobs, respectively, and the start-up subsidies – either affect every person in the labor force in the same way, or access to the program is open to every job seeker. That is, these measures constitute “universal treatments”, a fact that complicates the construction of a genuinely comparable counterfactual scenario, since the most straightforward scenario – “not affected by the policy measure” – does not exist. The identification of an appropriate counterfactual then requires meticulous construction of the comparison group. This is detailed for some measures in section 3.

In addition to the two conceptual problems, *two practical problems* also arise in creating the appropriate evaluation design for the Hartz reforms.

III. Implementation of the new steering system

Up to this moment – i.e. the point in time when the evaluation concept has to be set up – it remains unclear in what way exactly the steering system will be designed in detail, and to what extent the steering system has been implemented in the employment administration. Moreover, it is still an open question which information on the interaction between employment office staff and unemployed individuals will be collected and eventually made available to the evaluators.

With regard to the design of the evaluation this shortcoming implies that the concept can only suggest an ideal procedure, and the actual evaluation will –

where applicable – have to adjust or extend the set of indicators for the quality of labor offices correspondingly. As mentioned above, the importance of these indicators for a meaningful and comprehensive evaluation cannot be overemphasized: Without indicators for the quality of labor offices it is not possible to identify the causal effect of individual labor market policy measures – only the composite effect of single instruments together with the change in the general policy framework would then be identifiable.

IV. Data availability

Data availability is a crucial issue in setting up an evaluation concept. Currently, the data suitable for evaluation purposes provided by the employment administration (“Bundesagentur für Arbeit”) are about 12 months behind real time. This lag, as pointed out by the employment administration, can at best be reduced to 8 months. Since the first evaluation report will be due mid-2005, the data gap must be overcome by additional data collection.

In principle, all necessary data on socioeconomic characteristics and, most importantly, employment histories of unemployed individuals are available, and must then be complemented by data collection on *current* employment status and living situation. Moreover, depending on the specific labor market program being analyzed, additional data on program participation must be collected. Moreover, data on the indicators for the quality of labor offices (cf. above) must be added. Gathering these data will take several months, and subsequently the evaluator will need at least two months for analyzing the data. *Figure 1* depicts the resulting time frame for evaluating the Hartz reforms. The figure shows the starting date of several core measures of Hartz, along with the timing of sampling from the available data, additional data collection, and the evaluation.

Given the late starting date of some measures, the fact that program duration may be around 12 months, and the months needed subsequently for data collection and evaluation, it is clear that not all instruments of the Hartz reforms can be evaluated comprehensively until mid-2005 or mid-2006. Hence, the evaluation concept allows for a straightforward extension of the evaluation procedure beyond 2006, along the lines suggested in *Figure 1*, in order to assess long-term effects. The proposed time frame also has the advantage that, starting with 2006, each year the evaluation results of the previous year, which were then based on additional data collection, can be checked for consistency utilizing the now available register data from the employment administration (due to the 12 month time lag, cf. the note above).

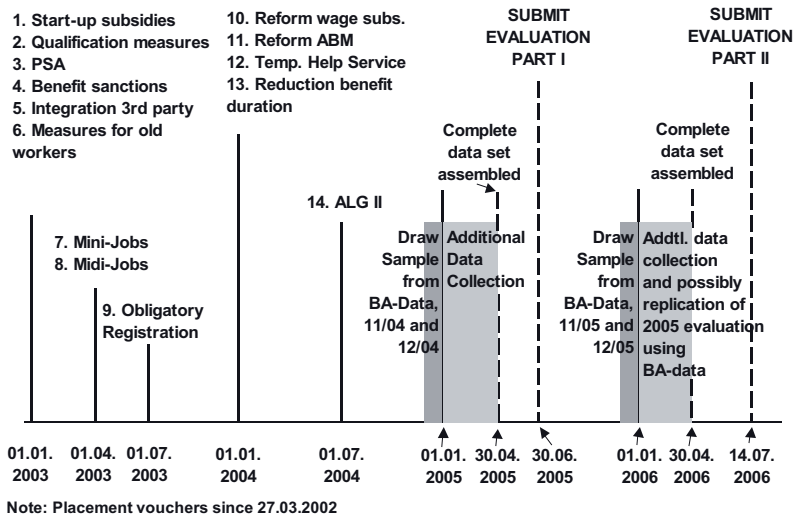


Figure 1: Time Frame of the Evaluation

3. Methodological Framework

The objective of any policy evaluation is the determination of the effects of the policy, its efficiency, and the reasons for success or failure. Specifically, these three steps are:

- I. Analysis of *effectiveness*: “What is the impact of the policy measure?”
- II. Analysis of *efficiency*: “What is the relation of benefits vs. costs of the measure?”, or: “At what cost has the effect been achieved?”
- III. Implementation and process analysis: “What are the reasons for the estimated effectiveness and efficiency?”

In principle this procedure holds for the evaluation on both micro and macro levels. Clearly, all three steps of the analysis are closely connected, and the omission of one of the steps would result in loss of important information for the decision maker. Specifically, simply focusing on step III and presuming that a measure must be effective if it is merely well-implemented (a practice certainly not uncommon in the past in Germany) would be quite naïve. Instead, the crucial step for generating robust empirical evidence is a cogent analysis of effectiveness. Steps II and III, while equally important, are logically subsequent, and aim at relating the established effect to its cost and possible reasons, respectively.

3.1 Descriptive Analysis

The first part of the empirical analysis of effectiveness concerns a descriptive account of the policy measure under scrutiny. This account provides the basis for the subsequent analysis. In the context of measures of labor market policy, the following elements are of particular interest:

- Description of objectives, core features, state of the implementation, and possible regional differences of the policy measure
- Assessment of incentive structures
- Description of characteristics and composition of program participants
- Gathering of information for identifying an appropriate comparison group, such as similarity in observable socioeconomic characteristics and employment histories

The descriptive analysis mainly aims at delineating and summarizing core statistical data that relate to program implementation, thus putting the evaluation into the appropriate context. This mainly serves a monitoring function.

3.2 Analysis of Effectiveness

The first, and most important, step in evaluating a policy measure regards the identification of the effect of the program. Estimating the causal effect of a program in general involves answering a so-called *counterfactual* question, i.e. making a statement about the hypothetical state of the world in the absence of the program. For instance, if we are interested in the effect of a qualification measure, and we observe an outcome of interest – such as employment probability – for program participants, then we would need to answer the counterfactual question: What would have happened to program participants in terms of their employment probability, if they had not participated in the program²? Whereas, obviously, program participants cannot be observed in two states of the world – with and without program participation – at the same time, answering the counterfactual is *the* methodological challenge of any program evaluation³. Without answering the counterfactual it is not possible to assess the impact of a policy measure, since it is exactly the difference between factual and counterfactual outcomes that measures the causal effect of the program.

² The evaluation parameter corresponding to this counterfactual is the so-called “Average Treatment Effect on the Treated”.

³ Hence, this problem is frequently referred to as the “evaluation problem”, cf., for instance, Heckman, LaLonde and Smith (1999), or Kluve and Schmidt (2002). The causal model underlying this counterfactual notion of causality is commonly labeled “potential outcome model”, since only one of the two outcomes required for causal inference is actually observable, and the second one is not, and thus a potential quantity. For further methodological details cf. Holland (1986) and Kluve (2004).

The fact that the counterfactual is not observable is called, in technical terms, *not identified*. It is therefore necessary to consider one or more *identification assumptions* that allow replacing the unobservable counterfactual with an observable counterpart. Ideally, the counterfactual would differ from the factual state in merely the fact that the policy measure was not implemented.

Besides these methodological challenges, several practical steps are important in analyzing program effectiveness. First, the unit of observation needs to be well-defined. In most cases concerning the Hartz reforms, unemployed individuals will be the units of observation of interest. In addition, for some measures, individual enterprises will constitute the appropriate units of observation. Second, the outcome of interest has to be singled out. Usually the outcome of interest corresponds to the objectives that the policy measure intends to achieve. It may be the case, however, that a program pursues several objectives at the same time, and that some of them may be competing objectives. In the case of the Hartz reforms, the most interesting outcome by far is employment, both since almost all particular measures aim at increasing individual participants' employment probability, and since the overall objective of the Hartz reforms is a reduction of the German unemployment rate.

Constructing the appropriate counterfactual, however, remains the most challenging task of the evaluation analysis. Given a modern economy with continuously growing production possibilities, hard-to-identify cyclical swings around a presumed growth path, and substantial heterogeneity in almost all economic aspects across individuals, sectors, and regions, it is clearly a difficult exercise to specify if some outcome of interest has taken on a "relatively high" value or not. Hence, finding a suitable comparison scenario determines success or failure of any impact analysis.

One important aspect in deciding on an identification strategy is individual heterogeneity of program participants. Quite clearly, program participants will differ from non-participants in a whole set of attributes, some of which will be observable, and others will be unobservable. Observable heterogeneity includes variables such as standard socioeconomic characteristics (age, gender, marital status, etc.) or individual employment histories. These characteristics are contained in the available data, and can therefore be utilized in the analysis, i.e. they can be controlled for when comparing the outcome of the group of participants (usually called "treatment group") with the outcome of a comparison group constructed from the pool of non-participants.

Unobserved heterogeneity, however, regards characteristics that are not contained in the data. We cannot observe whether an individual is characterized by e.g. motivation, stamina, or diligence. Unfortunately, these could be characteristics that partly determine the outcome of interest. If the treatment group systematically differs from the comparison group in one or more of these characteristics, then we would obtain biased estimates of the program effect. Hence, when choosing an

identification strategy it is a core challenge to take into account unobserved heterogeneity to the best possible extent. It then depends on the study design whether the strategy will be successful.

In general, an experimental study design is considered the “gold standard” and conceptually most convincing strategy for identifying causal effects. Since in an experimental study a population of individuals eligible for program participation is randomized into a treatment and control group, it can be shown that there will be no systematic differences – neither observable nor unobservable – between the two groups. Therefore the control group provides the counterfactual in a straightforward and conceptually conclusive way. Most evaluations of measures of active labor market policy, however, cannot rely on experimental data. Experimental evaluations in the labor market context have almost exclusively been put into practice in the US, with a very small number in Europe, and none of them in Germany. Likewise, the evaluation of the Hartz reforms relies on non-experimental data only⁴.

The statistical and econometrics literature discusses a large number of possible identification strategies based on different assumptions. All identification assumptions have in common that they are not statistically testable, i.e. they cannot be right or wrong *a priori*, or proven correct or false *a posteriori*; they can only be more or less plausible, or more or less easily violated. The literature contains many examples showing that, in analyzing the same program, different identification assumptions can lead to different impact estimates.

Choosing the appropriate evaluation strategy, i.e. identification strategy, for a specific program therefore involves the collection of relevant information that justifies the identification assumption. This information could come from details on e.g. program features, program participants, and program implementation, etc., but it generally requires detailed knowledge of program specifics. In the context of the Hartz reforms, one core variable in constructing comparison groups is the employment history. It has been frequently argued in the literature that such employment histories may capture possible differences between individuals due to unobserved heterogeneity to a large extent (Card and Sullivan 1988, Heckman and Smith 1999, Kluve, Lehmann and Schmidt 1999). Moreover, the appropriate evaluation strategy for the Hartz reforms relies on controlling for that type of observable heterogeneity originating in differences in the quality of services provided by the labor offices. Also the labor market context, specifically local or regional labor market conditions, constitute important control variables.

⁴ For an overview of evaluation studies of German labor market policy prior to the Hartz reforms cf. Fitzenberger and Speckesser (2000) and Fitzenberger and Hujer (2002). Kluve and Schmidt (2002) provide a review of the experience with Active Labor Market Policy in a European context.

The following identification strategies are commonly used in evaluation research:

A. Nonparametric methods

- Before-after
- Difference-in-Differences (DiD)
- Matching

B. Parametric methods

- Linear regression
- Fixed-effects panel model
- Duration analysis

C. Instrumental variables methods

- “Classic” IV method
- Arellano-Bond dynamic panel method

D. Discrete choice methods

In the context of the Hartz reforms, in the majority of cases non-parametric methods, in particular matching and DiD methods, will be the appropriate identification strategy. Some examples are given in section 4, where we discuss such suitable evaluation designs for specific instruments.

The basic idea of matching methods is to mimic a randomized experiment *ex post*. Utilizing information on a set of observable characteristics X , matching constructs – from a pool of potential comparison units – a retrospective comparison group as similar or comparable as possible to the treatment group in terms of X . The comparison group thus substitutes for the experimental control group. The main difference is that, whereas randomized assignment in an experiment balances both observable and unobservable attributes across treatment and control groups, matching can only control for observable covariates. The identification assumption, which matching is based on, is commonly referred to as “conditional independence assumption (CIA)” or “unconfoundedness”. Essentially it says that selection into treatment and comparison group is based on observables, and that, conditional on X , the non-participation outcome of the participant population (i.e. the desired counterfactual) can be replaced by the non-participation outcome of the non-participant population.

A related approach, the DiD method compares the difference in outcomes of participants and non-participants before and after the intervention. The difference of the two differences then measures the treatment effect. The approach is based on the identification assumption that unobserved heterogeneity between the two groups is intertemporally invariant; i.e. any differences in the outcomes due to unobserved characteristics do not change over time and will therefore be “differ-

enced out”. Further details of these and other identification strategies can be found in Heckman et al. (1999), and Blundell and Costas-Dias (2000); Fertig et al. (2004) discuss identification strategies in the context of the Hartz reforms at length.

3.3 Efficiency Analysis

Building on the impact analysis, the next step in evaluating a policy measure is to ask with which expenses the established effect has been achieved. This assessment of the benefits of the measure versus the costs of implementation is called *efficiency* analysis. If the preceding analysis of effectiveness has found a (qualitatively) positive effect of the intervention, then this is merely a necessary condition for a successful policy measure. In order to judge the success of the program, this effect needs to be confronted with the costs of the program. Clearly, this judgment is a relative measure, but most likely a program that causes only moderate improvement of the initial situation would not be allowed to be very cost intensive.

While consideration of the full costs of the program is desirable, the measurable *direct* costs – such as program fees or subsistence of participants – constitute only part of the total costs. Depending on the specific measure, also *indirect* costs have to be taken into account. For instance, opportunity costs may arise for program participants or firms: A program participant cannot exercise a job and effectuate earnings during program participation, and a firm may face additional administrative costs when managing a wage subsidies program. Moreover, it could be the case that a program has unintended negative side effects. For instance, a training measure could result in training participants displacing workers who did not participate in the program, or the training of participants would have been implemented even in the absence of government funding (cf. also Section 3.5).

Whereas, in principle, all direct costs can be taken into account in an efficiency analysis, this is much more difficult for indirect costs. Frequently, for instance, only coarse estimates for opportunity costs for workers and firms can be included. Unintended negative side effects can usually be assessed on an aggregate level, though methodological difficulties remain. However, consideration of program costs to the best possible extent is crucial for determining the success of a policy measure, as otherwise evaluation results might seriously misguide policy makers.

3.4 Implementation and Process Analysis

The logically last step of a full evaluation is an analysis of the implementation and the processes of a program. This step aims at identifying the *reasons* behind the effectiveness and efficiency established in the preceding stages. Necessarily, the implementation and process analysis proceeds much less formally, and focuses on qualitative rather than quantitative aspects of the program.

An important part of this analytical step is a discussion of the policy framework and context within which the individual employment offices act, along with the type and design of their cooperation with third parties. These two aspects constitute the frame for the *input* of resources and the provision of services as concrete activities for the clients. On the other hand, the actual services provided constitute the *output* for the clients.

In the empirical implementation of the evaluation, much of the data required for this part of the analysis will come from surveys of (a) local employment office employees, and (b) program participants or, more generally, registered unemployed individuals, i.e. all those eligible to receive services from the employment offices. Typical questions for this analysis are (for a full catalogue of questions for each policy instrument, cf. Fertig et al. 2004):

- From the perspective of local employment office employees, do possibilities for facilitating administrative procedures exist?
- Opinions on new / reformed instruments: Is the instrument (more easily) applicable?
- Are there any specific program features, specific efforts towards selecting or looking after particular workers, etc.?
- Reactions of program participants / of those affected by the policy measure.
- Analysis of data regarding indicators of labor office quality (cf. Section 2.3 above): Are there systematic differences in program effectiveness between employment office districts? If so, how do they relate to differences in the quality of service provision?

3.5 Distinct Features of the Macro Evaluation

In principle, the same methodological problems as on the micro level also arise for the evaluation on the macro level. The aim here is to identify the effect of the *entire* set of measures of labor market policy on an aggregate level. Again this requires answering the counterfactual question: “What would have happened to the outcome of interest (such as the unemployment rate) if the Hartz reforms had not been implemented?”

Similar to the micro level, appropriate identification assumptions are needed for construction of the counterfactual. Attention must be paid, though, to the potential endogeneity of policy measures: A labor market policy influenced by the current state of the labor market would imply that certain outcome variables (such as the unemployment rate or the share of long-term unemployed) determine the design of labor market policy, and not vice versa. Moreover, possible interactions between specific measures, and the interplay between labor market policy on a nationwide level and the levels of the federal states, need to be taken into account.

With respect to efficiency, the interest lies in identifying detrimental side effects that may potentially arise on the macro level (cf. Calmfors 1994):

- *Displacement effects*: Program participants take on jobs that, in the absence of the program, would have been taken on by non-participants.
- *Substitution effects*: Labor demand for other types of non-participants decreases due to changes in relative wages.
- *Deadweight loss*: The state funds programs that would have been implemented also in the absence of state funding.
- *Tax effects*: Financing of the Active Labor Market Policy through taxes has effects on non-participants.

In addition, the reduction of costs that may be brought about by a reduction in the number of unemployed needs to be estimated.

With respect to the implementation and process analysis, the following criteria should be taken into account (For a full catalogue of criteria cf. Fertig et al. 2004):

- How do local employment offices determine local labor market conditions, i.e. which information constitutes the basis for decision processes on spending of available funds?
- What is the influence of employment offices on the federal state level in this regard?
- Is there any cooperation or exchange of information between employment offices in the same region?
- Is there a local monitoring system, if so, does it entail essential features, and do monitoring results influence future implementations of measures?

Of particular importance are those criteria that influence the quality of service provision.

4. Evaluation Design

In this section, we delineate the salient practical components of an evaluation concept which adheres to the above mentioned guidelines and which is able to achieve the primary objective of generating *robust* and *comparable* evidence within the short time frame available for the concrete implementation of the evaluation. Clearly, the most challenging task in this endeavor is the generation of comparable evidence for the *effectiveness* of the different components of the Hartz reforms. Both the analyses of efficiency of the measures as well as the process and implementation analyses decisively depend on the results of the investigations regarding the measures' effectiveness. Therefore, we focus on the salient elements of the evaluation design aiming at this step of the overall evaluation strategy.

Table 2: Main Elements of Hartz Reforms – Suggestions for Analyses of Effectiveness

Counterfactual Question	Unit of Observation	Outcome Measure	Identification Strategy	Treatment Indicator / Comparison Group	Control Variables
<i>a) Public Job Provision Programs (ABM / SAM)¹⁾</i>					
What would have happened to the employment situation of participants, if they had not participated?	Individual unemployed	Employed (yes/no): 6 / 12 months after end of program	Matching on observables (requires CIA to hold)	Participation in ABM / SAM; comparison group: untreated unemployed ²⁾	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
<i>b) Qualification Measures (Fbw)¹⁾</i>					
What would have happened to the employment situation of participants, if they had not participated?	Individual unemployed	Employed (yes/no): 6 / 12 months after end of program	Matching on observables (requires CIA to hold)	Participation in Fbw; comparison group: untreated unemployed ²⁾	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
What would have happened to the employment situation of participants, if the reform of Fbw had not happened?	Individual unemployed	Employed (yes/no): 6 / 12 months after end of program	Difference-in-differences (assuming inter-temporal invariance of unobserved heterogeneity)	Participation in Fbw before and after the reform; comparison group: untreated unemployed ²⁾ before and after the reform	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation

Table 2 (continued)

Counterfactual Question	Unit of Observation	Outcome Measure	Identification Strategy	Treatment Indicator/ Comparison Group	Control Variables
<i>c) Direct Wage Subsidies (EGZ)1)</i>					
What would have happened to the employment situation of participants, if they had not participated?	Individual unemployed	Employed (yes/no): 6/12 months after end of subsidy	Matching on observables (requires CIA to hold)	Supported by EGZ; comparison group: untreated unemployed ²⁾	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
What would have happened to the employment situation of participants, if the reform of FbW had not happened?	Individual unemployed	Employed (yes/no): 6/12 months after end of subsidy	Difference-in-differences (assuming inter-temporal invariance of unobserved heterogeneity)	Supported by EGZ before and after the reform; comparison group: untreated unemployed ²⁾ before and after the reform	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
<i>d) Personnel Service Agency (PSA)1)</i>					
What would have happened to the employment situation of employees in PSAs, if they had not been employed there?	Individual unemployed	Regularly employed (yes/no) during or directly after PSA-period	Matching on observables (requires CIA to hold)	Employment in PSA; comparison group: untreated unemployed ²⁾	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation

Table 2 (continued)

Counterfactual Question	Unit of Observation	Outcome Measure	Identification Strategy	Treatment Indicator/ Comparison Group	Control Variables
e) Self-Employment Start-Up Subsidies (Ich-AG and ÜG1)					
What would have happened to the employment situation of participants, if they had not participated?	Individual unemployed	Employed (yes /no): 6 / 12 months after end of subsidy	Matching on observations (requires CIA to hold)	Supported by Ich-AG or ÜG; comparison group: untreated unemployed ²⁾	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
What would have happened to the employment situation of participants, if they had received alternative funding?	Individual unemployed	Employed (yes /no): 6 / 12 months after end of subsidy	Matching on observations (requires CIA to hold)	Supported by Ich-AG or ÜG; comparison group: self-employed receiving Start-Geld or Micro-Darlehen from KfW	Socio-economic characteristics; (un-) employment history; indicators for regional labor market situation
f) Mini- and Midi-Jobs					
What would have happened to the employment situation of unemployed taking up a Mini- or Midi-Job, if they had not done this?	Individual unemployed	Regularly employed (yes /no): 6 / 12 months after start of Mini- or Midi-Job	Matching on observations (requires CIA to hold)	Unemployed taking up a Mini- or Midi-Job; comparison group: untreated unemployed ²⁾ without Mini- or Midi-Job	Socio-economic characteristics; (un-) employment history; indicators for regional labor market situation
What would have happened to the employment situation of unemployed taking up a Mini- or Midi-Job, if the reform of these jobs had not happened?	Individual unemployed	Regularly employed (yes /no): 6 / 12 months after start of Mini- or Midi-Job	Difference-in-differences (assuming inter-temporal invariance of unobserved heterogeneity)	Unemployed taking up a Mini- or Midi-Job before and after the reform; comparison group: untreated unemployed ²⁾ without Mini- or Midi-Job before and after the reform	Socio-economic characteristics; (un-) employment history; indicators for regional labor market situation

Table 2 (continued)

Counterfactual Question	Unit of Observation	Outcome Measure	Identification Strategy	Treatment Indicator/Comparison Group	Control Variables
What would have happened to the number of Mini- or Midi-Jobs, if the reform of these jobs had not happened?	Individual companies	Number of employees in Mini- or Midi-Job	Regression model for period before reform; ass. structural stability of estimates and evaluating post-reform covariates with them yields predicted jobs after reform; difference between actual and predicted number of jobs yields effect of reform	Comparison group: all other companies in the sample	Company-specific characteristics; indicators for regional labor market situation
g) Merging of Unemployment Assistance and Social Assistance (ALG II)					
What would have happened to the employment situation of ALG II-recipients, if they had received unemployment assistance instead?	Individual unemployed	Employed (yes/no): 5 months after start of benefit receipt	Matching on observables (requires CIA to hold); utilizing transitional regulation	Receipt of ALG II and entitled for unemployment assistance in old regulation context; comparison group: receipt of unemployment assistance directly before the reform came into effect	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
What would have happened to the employment situation of ALG II-recipients, if they had received social assistance instead?	Individual unemployed	Employed (yes/no) on a monthly basis	Before-after comparison in duration analysis framework	Receipt of social assistance or ALG II	Key socio-economic characteristics (family situation, age, education); duration of spell

Notes: 1) Direct comparisons (using matching techniques) between these instruments of ALMP recommended. 2) Untreated unemployed denotes individuals receiving no special treatment, i.e. individuals not participating in any measure of ALMP.

Table 2 provides an overview of the most important measures of the Hartz reforms, and details suggestions for the evaluation of their effectiveness⁵. Specifically, the table presents relevant counterfactual questions and outlines the main ingredients of an identification strategy able to provide an answer to these questions. We would like to emphasize that direct comparisons between some of the instruments mentioned in Table 2 are recommendable, in order to assess which of the respective instruments is relatively more successful in promoting the job prospects of unemployed individuals.

To render such an approach feasible, a unified data collection and estimation strategy is indispensable. The cornerstone of our evaluation concept is therefore a unified scheme for the collection of individual data on participants in a particular program and their respective comparison group. Regarding the implementation of our concept, this crucial element implies that only *one* comparison group needs to be constructed that can be used simultaneously for the evaluation of the majority of instruments. Not only does this setup reduce the effort as well as the costs for the collection of data, but it also ensures comparability and unified interpretability of evaluation results across instruments.

In order to estimate the mean effect of treatment on the treated a variety of individual-level data for participants and non-participants as well as additional information is necessary. Specifically, we need *individual employee data*, especially the current labor market status and the employment history, as well as socio-economic characteristics and information on family background (labor market and earnings situation of partner and children). The labor market history as well as the family background information of the individuals are of particular importance, since this information is decisive for alleviating the potential problem of unobserved heterogeneity between participant and non-participant groups (see also section 3).

Furthermore, for some measures (especially the evaluation of Mini- and Midi-Jobs) it is necessary to collect *individual employer data*. That is, we need information on the number of employees (disaggregated by their social insurance coverage) as well as on company characteristics like the year of foundation, main sector of business, investment activities, organizational change and others.

Finally, data on the level of the local labor offices needs to be collected. Specifically, it is indispensable to collect indicators for the quality of service provision in the 180 local labor offices. Without such indicators the impact of the various components of the Hartz reforms cannot be identified, and cannot be isolated from the effect of the changes in the administration and steering of labor market policy.

⁵ As pointed out various times throughout the paper, the Hartz reforms comprise a multitude of particular policy changes and policy interventions. For the sake of brevity, in this section we focus on the main elements in the whole set of reforms. For further details on secondary measures cf. Fertig et al. (2004).

The primary data source for providing this information is the *Federal Employment Agency*. However, due to the time lag in data processing or partially incomplete data for evaluation purposes (see Section 2), it is necessary to complement the information at the *Federal Employment Agency* by additional data collection. *Figure 2* and *Figure 3* illustrate the unified data collection scheme for both evaluation reports due 2005 and 2006, respectively.

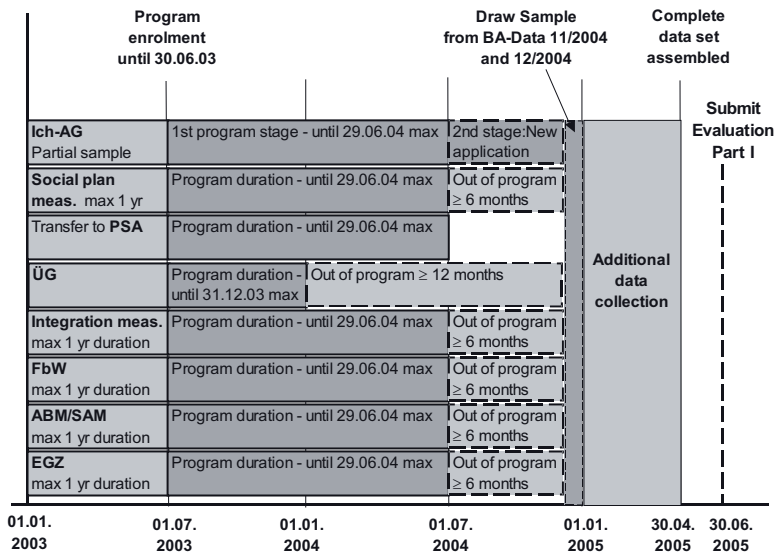


Figure 2: Uniform Sampling Concept, short-term

The idea behind this scheme is to draw random samples of participants for the various measures, who enrolled in these measures between January and June 2003 and 2004, respectively, from the database of the *Federal Employment Agency*. For these participant groups only one comparison group of untreated unemployed has to be drawn⁶. The participant groups as well as the common comparison group then have to be interviewed regarding their labor market status during 2004 and 2005, respectively, as well as with respect to their family background⁷. The collected information has to be matched to the existing data at the BA to construct the samples necessary for the empirical investigation of program effectiveness. Clearly, since all participants entered the different measures within the same

⁶ “Untreated unemployed” denotes registered unemployed individuals who did not participate in any measure of ALMP.

⁷ For cost reasons our concept recommends doing this by sending out questionnaires, since the number of individuals to be interviewed is rather large.

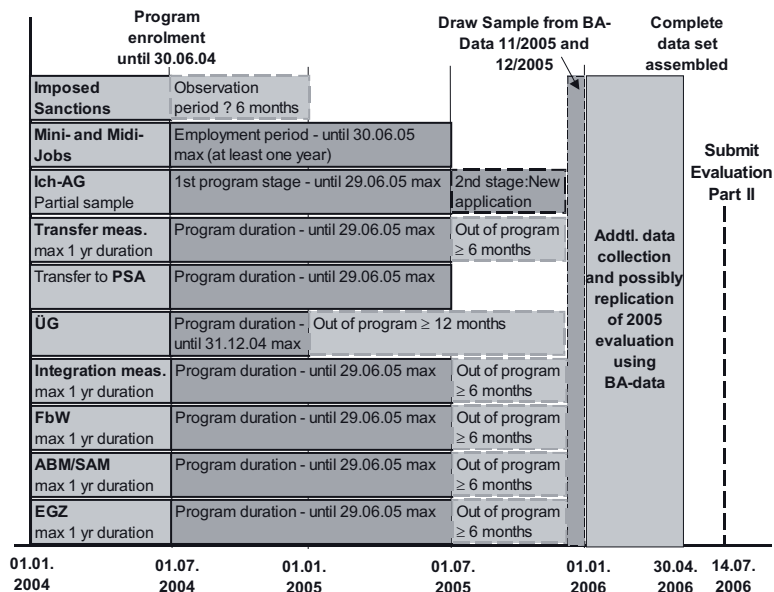


Figure 3: Uniform Sampling Concept, medium term

period, it is possible to compare the labor market outcomes of different participant groups directly, in order to investigate which measure is relatively more successful in increasing the labor market prospects of the unemployed. Finally, for a small set of components of the Hartz reforms, especially those regarding benefit sanctions, it is necessary to construct additional comparison groups. However, this can be accomplished with comparably little effort.

5. Conclusion

It seems likely that the Hartz reforms are among the most ambitious, comprehensive, and far-reaching policy reforms in Germany over the last decades. Aiming at a sustainable contribution to reducing unemployment, the Hartz reforms entail a multitude of specific labor market policy changes, introduction of new instruments, and modifications of the general administrative framework in which German labor market policy operates. It is, however, anything but clear *a priori* that this bold endeavor will attain its objectives.

Learning about the success, predominantly the effectiveness, of any policy measure requires systematic evaluation. Given the manifoldness and scope of the Hartz reforms, this is a particularly challenging task. Fortunately, German policy makers have decided to rely on academic expertise in this regard, and to include academic

experts from the very outset of evaluating the reforms, i.e. setting up an appropriate evaluation concept. This creates a novel situation for German evaluation practice in two regards: For the first time academic experts were consulted prior to the implementation of a policy, which in turn will result in the first systematic and comprehensive policy evaluation in Germany. The significance of this development cannot be overemphasized.

As the Hartz reforms are currently being set into practice, the task with which the government approached academics can be summarized as follows: to develop a ready-to-implement concept for the evaluation of the full set of reforms in their entirety, as well as each instrument on its own. The main restriction regarding this task is the time constraint: The concept needs to allow for generating of robust evidence until mid-2005 (1st report) and mid-2006 (Final report), respectively.

In this paper, we have outlined such a concept. At the outset, the paper has discussed elements and objectives of the Hartz reforms, and has detailed essential guidelines for an appropriate evaluation framework. While, in principle, outlining an ideal evaluation concept, we have discussed limitations that arise for this concept both in practice and from a conceptual point of view. More importantly, facing these limitations, we have discussed ways to overcome practical and conceptual obstacles.

The paper has proceeded to discuss central methodological issues regarding the three main steps of an evaluation study, i.e. analyses of effectiveness, efficiency, and implementation and process analysis. Finally, we have presented the core features of a suitable evaluation design for the Hartz reforms, mainly focusing on tailor made evaluation approaches for specific policy measures, and a sampling scheme for the data collection that is both conceptually appropriate and practicable.

In summary, the evaluation design we suggest takes into account the contextual specifics of the Hartz reforms, provides feasible solutions to conceptual and practical problems, and will enable the evaluator to generate the desired empirical evidence given substantive time and data restrictions. In addition to the fact that our concept is directly implementable, it also has the advantage of being extensible in a straightforward way, once the data basis has been set up. Hence, we are confident that this concept for evaluating the Hartz reforms can contribute substantially to a cogent, comprehensive, and lasting evaluation of labor market policy in Germany in the long run.

References

- Blundell, R. / Costas-Dias, M.* (2000): "Evaluation Methods for Non-experimental Data", *Fiscal Studies* 21, 427–468.
- Calmfors, L.* (1994): "Active Labour Market Policy and Unemployment – A Framework for the Analysis of Crucial Design Features", *OECD Economic Studies* 22.
- Card, D. / Sullivan, D.* (1988): "Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment", *Econometrica* 56, 497–530.
- Fertig, M. / Kluve, J. / Schmidt, C. M. / Friedrich, W. / Apel, H. / Hägele, H.* (2004): "Die Hartz-Gesetze zur Arbeitsmarktpolitik – Ein umfassendes Evaluationskonzept", *RWI Schriften* 74, Duncker & Humblot, Berlin, forthcoming.
- Fitzenberger, B. / Hujer, R.* (2002): "Stand und Perspektiven der Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland", *Perspektiven der Wirtschaftspolitik* 3, 139–158.
- Fitzenberger, B. / Speckesser, S.* (2000): "Zur wissenschaftlichen Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland: Ein Überblick", *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 33, 357–370.
- Heckman, J. J. / LaLonde, R. J. / Smith, J. A.* (1999): "The Economics and Econometrics of Active Labor Market Programs", in *Handbook of Labor Economics*, 3, edited by O. Ashenfelter and D. Card, Elsevier, Amsterdam.
- Heckman, J. J. / Smith, J. A.* (1999): "The Pre-program Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies", *Economic Journal* 109, 313–348.
- Holland, P. W.* (1986): "Statistics and Causal Inference" (with discussion), *Journal of the American Statistical Association* 81, 945–970.
- Kluve, J.* (2004): "On the Role of Counterfactuals in Inferring Causal Effects", *Foundations of Science* 9, 65–101.
- Kluve, J. / Lehmann, H. / Schmidt, C. M.* (1999): "Active Labour Market Policies in Poland: Human Capital Enhancement, Stigmatization, or Benefit Churning", *Journal of Comparative Economics* 27, 61–89.
- Kluve, J. / Schmidt, C. M.* (2002): "Can training and employment subsidies combat European unemployment?", *Economic Policy* 35, 409–448.

A Conceptual Framework for the Evaluation of Comprehensive Labor Market Policy Reforms in Germany

Comment

By Ronnie Schöb*

There is promising news: the German Bundestag is explicitly demanding the evaluation of the new labor market reform measures, the so-called Hartz reforms, on a solid scientific basis. For the first time in Germany, economists are involved from the very beginning in the implementation and evaluation of current labor market reforms. The downside is that politics requires results within a very short time period. A first report is expected by mid-2005 and a final report by mid-2006. This is a very short time horizon and a huge challenge for evaluation experts.

In their paper, Fertig and Kluve accept this challenge and present a comprehensive concept to evaluate the full set of the Hartz reform instruments as well as each instrument on its own. Their aim is, despite the tight time schedule, to come up with reliable empirical evidence by the end of the current legislative period. Three questions are mainly dealt with in this paper:

1. How can we assess the effectiveness of labor market reforms?
2. How can we assess the efficiency of labor market reforms?
3. How can we determine the reasons for the estimated effectiveness and efficiency of the reforms?

To address these questions, Fertig and Kluve discuss a comprehensive evaluation design, specific methods applicable to particular reforms, and ways in which the required data can be collected and analyzed. At length they discuss the pros and cons of their evaluation concept, the scope and limitations of their approach, and finally present a ready-to-implement strategy for evaluating these reform proposals that is very much state of the art.

In my comment, I will address two questions that I consider essential when it comes to the scientific evaluation of current policy measures. First, economists

* Otto-von-Guericke-University Magdeburg and CESifo, Munich. Address: Otto-von-Guericke-University Magdeburg, Department of Economics and Management, P.O. Box 4120, D-39016 Magdeburg, Germany, Email: ronnie.schoeb@www.uni-magdeburg.de

should emphasize the importance of a consistent theoretical economic model on which any reasonable evaluation must be founded. Any econometric evaluation must refer to an economic model in which the mechanisms of the labor market reform under consideration are made transparent and where possible shortcomings are identified. Second, economists should not be too modest when it comes to scientific evaluation of social policy. Simply comparing reforms to the “taking-no-action future” may be completely misleading. As our profession has made tremendous progress in developing comprehensive reform proposals for the German labor market based on thorough economic analysis, we should not be too modest when it comes to evaluating policies. Instead of comparing the Hartz reform proposal with the worst-case scenario, the status quo carried forward into the future, the current reforms should also be compared with promising alternative reform proposals. This would allow us to rank the current labor market reforms on an interval scale rather than comparing them with the taking-no-action scenario only. Of course, for such a comparison, no empirical data is available. But we should not restrict ourselves to searching for the key in the places that the street lamps light up.

Theory first!

Whether we have to deal with a single labor policy measure or with a bundle of measures, as in the Hartz reforms, we definitely need a thorough theoretical economic analysis of these measures before we can start to evaluate them. Such a theoretical analysis requires a clear statement of objectives, a detailed description of the instruments under consideration, and the economic and institutional environment in which these instruments are implemented. The final objective of such an analysis is to predict the outcome of the Hartz reforms, to identify the main mechanisms at work, and to show where the reforms might work and where not.

Obviously, the main objective of the Hartz reforms is a significant reduction of unemployment in Germany. For evaluating the reforms, however, this ultimate goal is split in intermediate objectives such as a quicker and more sustainable job placement and the creation of new jobs and employment opportunities. This seems a straightforward exercise to do but it is not at all obvious that either of these intermediate objectives will help achieve the main objective of alleviating unemployment. Two examples should suffice to show the potential traps of such a procedure.

First, consider the implementation of the Personal Service Agencies. Their objective is to bring unemployed workers back into employment by providing temporary help service jobs that should, in the end, allow them to find a new permanent job. The suggested evaluation will mainly focus on job placement and may (or may not) demonstrate the effectiveness of Personal Service Agencies with respect to the intermediate objective “job placement”. But the way in which these jobs are subsidized crowds out regular temporary work. This impact is not consid-

ered in usual evaluations. An evaluation may therefore indicate success where no additional jobs were actually created¹: a cost-benefit analysis with respect to job placement for the formerly unemployed may indicate that this intermediate objective is reached more efficiently than with previous instruments. With respect to the ultimate goal, however, the judgment would be that there is no benefit at all but at immense cost! A thorough *a priori* theoretical foundation would help circumvent such an evaluation trap.

The problems become even more pronounced when we look at a second example, the new Mini- and Midi-Jobs implemented last year. For this measure, Fertig and Kluve list several objectives, of which we will discuss two only. One objective is the integration of the unemployed by providing incentives to take up such jobs that are supposed to become stepping stones into regular work. Another one is to increase the flexibility of employment in low-paid jobs. While the second goal will certainly be reached, economic theory predicts that the first goal won't. First of all, most of the additional Mini-Jobs are just substitutes for former regular full-time jobs in the service sector and do not actually raise employment. Apart from this effect, which will hardly be detected in an evaluation study focusing on the unemployed only, Mini-Jobs will also create new jobs and this effect will become visible in the data. But the evaluation of the Mini-Job regulation with respect to the intermediate objective "job creation" may be flawed, as it does not take into account that these regulations mainly benefit those interested in a second part-time job, for secondary wage earners, students and pensioners. Although these regulations reduce income taxes and social security contributions and thus narrow the wedge between net-of-tax wage and labor cost, they do not reduce the implicit tax of up to 100% unemployment benefit recipients have to bear when they take up work. The Mini-job regulations have created a low-paid sector where low-skilled unemployed have to compete against other job seekers. Increasing employment in this sector does not go along with lower unemployment – quite the reverse is likely to occur.

Less modesty!

Evaluating labor market policies requires a reference scenario. Ideally, such a reference scenario can be constructed in experimental studies where one can separate *ex ante* the individuals eligible for a particular labor program into a treatment and control group without creating systematic differences between these groups. If this is not possible, one has to compare the outcome with the hypothetical state of the world in the absence of the program. This can be done with a "potential outcome model" that tries to identify a suitable control group and compares the outcome of this group with the treatment group. As is shown in the

¹ For a detailed description of the crowding out problems of Personal Service Agencies, see Schöb and Weimann (2003, pp. 108 – 114)

paper by Fertig and Kluve, there are sophisticated methods available to design such a model.

But even the most sophisticated methods do not overcome the problem that the only comparison we have is the future outcome that occurs if the program were not to have been implemented. The whole approach is not designed for a comparison of the program with alternative proposals. In principle, there is nothing wrong with such a limited comparison. But one has to be explicit in stating that one is analyzing the achievements relative to a status quo carried forward into the future and not relative to some potentially superior program. In other words: without being aware of this limitation one is easily inclined to mark a third-best measure as a success because it allowed for modest improvement although it shows disappointing results compared to some better-suited measure.

Empirical evaluation on its own, cannot overcome this limitation. Additional theoretical analysis is needed. And here, less modesty is required. There are many serious proposals for comprehensive labor market reforms, such as the proposals of the German Council of Economic Experts, the Scientific Council at the Federal Ministry of Economics and Labour, the Ifo Institute as well as the proposal of wage subsidies for employers, called the “Magdeburg Alternative”.² All these proposals aim at a fundamental change in the way the German welfare system is organized by moving from welfare to workfare. Comparing the Hartz reforms with these proposals may yield completely different results to the comparison with the status quo. There is the demand for evaluation but there is one degree of freedom: the choice of the reference scenario. Economists should not hesitate to compare current policy with the best policy available. The final report due mid-2006 should not only report progress made but also state potentials forgone.

References

- Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung* (2002): “Jahresgutachten 2002/2003. Zwanzig Punkte für mehr Beschäftigung und Wachstum”, Wiesbaden, <http://www.sachverstaendigenrat-wirtschaft.de/gutacht/02ges.pdf>.
- Schöb, R./ Weimann, J. (2003): “Arbeit ist machbar. Die neue Beschäftigungsformel”, Stekovics, Döfel, www.arbeitistmachbar.de.
- Sinn, H.-W. et al. (2002): “Aktivierende Sozialhilfe: Ein Weg zu mehr Beschäftigung und Wachstum”, ifo-Schnelldienst 55(9), <http://www.ifo.de/link/SD09-02TextS03.pdf>.
- Wissenschaftlicher Beirat beim Bundesministerium für Wirtschaft* (2002): “Reform des Sozialstaats für mehr Beschäftigung im Bereich gering qualifizierter Arbeit”, BMWa-Dokumentation 512, August 2002, <http://www.bmwi.de/Navigation/arbeit,did=6458.html>.

² See Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2002), Wissenschaftlicher Beirat beim Bundesministerium für Wirtschaft (2002), Sinn et al. (2002), Schöb and Weimann (2003).

Health-Care Efficiency in OECD Countries

By Rigmar Osterkamp*

Abstract

Efficiency of health-care provision is measured for 22 European and seven non-European countries in 1980, 1990, and 2000 by means of the non-parametric method of Free Disposable Hull (FDH) analysis. As output variable, potential life years lost (reciprocal value) is used, as input variables health expenditure and health employment. It appeared necessary to correct the input variables for inter-country differences of medical incomes and age structure. It turns out that (a) there is a considerable difference between the efficiency measures for uncorrected and for corrected input variables, (b) the ranking of efficiency puts Japan, Spain, Sweden, Greece and Portugal on top, while Finland, France, Germany, Slovak Republic, Switzerland and USA occupy the lower spots, and (c) efficiency has improved considerably between 1980 and 2000 in Italy and Sweden, while it has deteriorated in the USA. The question of the reasons for the efficiency rankings and for the changes in efficiency is not posed in this paper.

JEL classifications: I12, I18, H51, C14

Keywords: health-care efficiency, production possibility frontier, FDH analysis, absenteeism, lifestyles, comfort, medical incomes, age structure

1. Introduction

The necessity of economic reforms in Europe applies not only to the much-discussed sectors of labor market, capital market, public enterprises, public finances and pensions. Health-care consumes about a tenth of GDP: thus, whether this sector is organized in an efficient or inefficient way makes a difference.

* Ifo Institute for Economic Research, Munich.

I am indebted to the article by António Afonso and Miguel St. Aubyn (2003), which provided many ideas and suggestions for the line of research I have pursued here. For a stimulating critical discussion of this paper and of alternative methods to estimate efficiency of health-care systems, I am indebted to the participants in the ARGE conference, but specifically to Peter Zweifel, University of Zurich, and for the remarks he made in his Comment. Moreover, I thank the editor of AEQ for valuable additional remarks. For errors, omissions and mistakes of this paper, however, I take the sole responsibility.

Twenty-two European and seven non-European countries have been analyzed here for the years 1980, 1990, and 2000. The non-parametric method of Free Disposable Hull (FDH) analysis is applied¹. This method provides a relative measure of efficiency in the sense that a country's efficiency measure depends on the efficiency of one or more other countries. That is, a country may fall back in terms of the measured FDH efficiency solely because another country has improved its efficiency.

Health-care efficiency in a country-comparative perspective is seemingly still a rare subject in the health economics literature. Even in the rather comprehensive "Handbook of Health Economics" (Culyer and Newhouse 2001), the term "efficiency" does not show up. "Productivity" is treated there, but only some special aspects of it (in relation to alcohol, waiting lists, dental services). The article on "International Comparisons of Health Expenditure" (Gerdtham and Jönsson) in the same book also does not deal with efficiency questions. However, in a recent article, Afonso and Aubyn (2003) measure the efficiency of health-care expenditure (as well as public education spending) in OECD countries – but do not consider its development over time or correct the input variables for certain factors.

The paper is organized as follows. In Chapter 2 we will look briefly at the method employed here to measure efficiency. In the next part (Chapter 3) the selection of the output and the input variables is explained. Chapter 4 discusses the necessity to correct the input variables for those differences between countries that should not influence the efficiency measure. There are several such differences, but due to data limitations, only differences of medical incomes and age structure can be allowed for. Chapter 5 presents the results. Chapter 6 summarizes the results, and Chapter 7 formulates some questions for further research.

Annex 1 (Tables 1 to 4) contains the result of the FDH calculations. Annex 2 (Tables 5 to 8) presents the rankings and the changes of efficiency, derived from the FDH calculations. Annex 3 (Table 9) contains the raw data. Annex 4 presents a short introduction to the application of FDH analysis.

2. A Brief Look at FDH Analysis

FDH analysis allows numerical measures of efficiency to be attached to each country². Efficiency is expressed in two measures, as input efficiency and as output efficiency. Both measures are relative measures of efficiency in the sense that the efficiency of a "non-dominant" country is compared to that of a "dominant" country. The efficiency scores of the latter are, by definition, 1. The efficiency scores of the non-dominant countries are lower than 1, but always positive. A non-

¹ For an assessment of alternative methods to estimate health-care efficiency see the Comment by Peter Zweifel to this paper.

² More generally, to each "decision unit" as they are called in the FDH literature.

dominant country is one that produces less output with more input than one or more other countries, while a dominant country is one for whom there does not exist any other country that produces more output and uses less input. The dominant countries form the “production possibility frontier”, or more simply the “production frontier” or “frontier”. The efficiency score of the non-dominant countries depends on their “distance” to the production frontier. The larger the distance, the smaller is the efficiency score. FDH analysis prescribes a specific algorithm to measure that distance³. The distance is expressed output-wise (output efficiency score) as well as input-wise (input efficiency score).

FDH analysis can handle, principally, any number of inputs and outputs. However, the higher the sum of outputs and inputs is, the higher the number of dominant countries and the lower the number of non-dominant countries left for the calculation of their distance to the frontier. Moreover, limitations in the availability of data – specifically for some important variables as well as for historical values – that reduce the number of countries to be possibly included in the analysis led to the restriction in this paper to use only one output and only one input in each of the calculations.

The Annex contains an explanation of how to apply FDH analysis.

3. Selection of Variables and Years

Output variable

For an efficiency analysis of health-care systems, some output variables lend themselves immediately as candidates: infant mortality, probability of dying between 15 and 59 years, life expectancy at birth or at various ages, healthy life expectancy, potential life years lost due to premature mortality.

Infant mortality and the probability of dying between 15 and 59 years are important socio-economic concepts, but are of a partial nature. The other variables are more comprehensive. Potential life years lost due to premature mortality (OECD Health Data) has the advantage that also historical data are available. Thus, this variable (its reciprocal value) was chosen for the further analysis.

Input variable

A variable in monetary terms, such as total health expenditure, has the advantage of being a comprehensive measure. It must be related, of course, to either GDP or to the population. The relation to GDP – instead of to the population – has the advantage that the measure does not increase when incomes in the medical sector rise with – and not more than – per capita income. The disadvantage of such

³ For a useful introduction to FDH analysis with practical examples see, e.g., Afonso and Aubyn (2003) or Gupta and Verhoeven (2001).

a variable is that it is influenced by factors that are not under the control of public health authorities, and thus would distort efficiency analysis. Insofar as this is the case, the variable should be corrected. This is discussed in the following chapter.

A variable in non-monetary terms, like total health employment, has the advantage that it is directly related to health care and (more or less) under the control of health-care authorities. But the disadvantage is that such a non-monetary variable can never be a comprehensive measure of the input of the health-care system.

This paper will use the comprehensive monetary variable “total health expenditure” (OECD Health Data) as well as the partial non-monetary variable “total health employment” (OECD Health Data) in separate approaches, i.e., in each approach only one input variable is used.

Years

Reforms of the institutional design of health care systems and changes of their efficiency occur, in most cases, in small steps, although there are also examples of reform shocks. Thus, it seems to be appropriate to show the development of efficiency over time not on an annual basis but from decade to decade. Unfortunately, the most comprehensive health data – those of the database “OECD Health Data” – exhibit for many important variables a rising share of loopholes the farther one goes back. 1980 is the earliest year for which an efficiency analysis can reasonably be made. Thus, we consider the development of efficiency by looking at 1980, 1990, and 2000.

4. Correction of the Variables

The input variables health expenditure and partly also health employment are influenced by at least six factors that might distort an efficiency analysis: absenteeism due to sickness, lifestyles, comfort, incomes in the medical sector, age structure and climate (environment). Should the input variables named therefore be corrected for the influence of these factors? Is such a correction possible, and if so, how can it be done? These questions are exemplified for the monetary variable of costs of health expenditure.

Absenteeism due to sickness

A variable such as “total costs of health care” does not seem to exist in databases. What exists is only “total expenditure on health” (e.g., in OECD Health Data). One major difference between total costs and total expenditure is the loss of production that results from absenteeism from the workplace due to sickness. Absenteeism due to sickness varies largely over countries, while the variations seem to stem to a large degree from different incentive structures (generosity of granting sick leave). It is debatable how the loss of production due to sickness-related absenteeism from the workplace should be taken into consideration in measuring

health-care efficiency. On the one hand, “absenteeism due to sickness” is surely an output (or outcome) variable, but on the other it is also an input (cost) variable. Unfortunately, lack of data, specifically for earlier years, does not allow the introduction of absenteeism – in whatever form – as a correction factor⁴.

Lifestyles

Lifestyles differ between countries and change over time. Obesity, smoking behavior, alcohol consumption and – positively – physical exercise, for example, are surely relevant to the costs of health care. But should the impact of these variables be accounted for? It could be done, in principle because the data availability is relatively good. For the case of obesity in the U.S. it *has* also been done (Comanor 2003)⁵.

The effect of corrections for – good or bad – lifestyles would be that countries with good (bad) lifestyles of their populations would receive worse (better) efficiency scores than they would receive without such a correction. In this paper, I take the position that the health-related quality of lifestyles is an important duty and – at least partially – an important effect of health care provision, of public health activities and specifically of health education. Thus, there will be no correction for different lifestyles in this paper.

Comfort

In most countries, health-care services are seldom provided in a pure way. Instead, they are generally accompanied by an element of comfort. A lower number of beds per room in a clinic, a higher number of nurses for a given number of patients, a higher regional density of practicing physicians – all this might (or might not) contribute to better health care proper and might (or might not) improve the output variable in efficiency analysis, but provides also more comfort to the patients (perhaps also to the health care providers) and increases that total which makes up “total expenditure on health”. A certain amount of comfort seems to be an unavoidable element of health-care services. Health-care services proper and comfort proper can be regarded as joint products or as unavoidable by-products⁶.

⁴ Calculations to grasp this factor have been done for the period of 1998–2000 in Osterkamp (2002, a).

⁵ The author comes to the conclusion that, by fully taking into account the extra-costs of health care due to obesity, total health expenditure in percent of GDP in the U.S. would not be much higher than in most other countries of comparable per capita income.

⁶ This reasoning might even lead to questioning whether health or health-care are superior goods, which is – since the seminal work of Newhouse (1977) – a widely-shared suggestion of health economists. However, the fact that health care expenditure rises over-proportionately with rising incomes might be – at least partially – due to an income elasticity higher than 1 for *comfortable health-care services*, not so much for health or health-care services proper.

If the element of comfort in health care provision differs between countries and if it changes over time – both seems to be obvious – and if comfort which accompanies health services is not identically related to per capita income (probable, but not obvious), then country-comparative *health care-efficiency* measures which are not corrected for different amounts of comfort might be distorted. Thus, such a correction should be done. Comfort is, similarly to absenteeism, on the one hand an output (outcome) variable (there is a demand for comfort) but on the other, an input (cost) variable. In contrast to absenteeism due to sickness, the costs of comfort are automatically included in “total health expenditures”.

Unfortunately, the joint product characteristic of health care proper and comfort proper makes a separation of the two elements difficult, if not impossible. At least, there does not seem to be a variable that lends itself easily to such a correction. A proxy could perhaps be the average number of beds per room in hospitals because a higher number of beds per room might primarily result in lower comfort and not so much in reduced health care proper. Unfortunately, for such a variable no data is available to my knowledge. As another candidate for a proxy one could think of the number of health personnel per 1000 inhabitants. This variable surely entails an element of comfort, but seems to be much more related to health than to comfort. Thus, the paper has to proceed without such a correction, although a correction would be necessary for a proper assessment of efficiency of health-care services.

Climate

It is quite plausible that climatic conditions also influence health outcomes and, thus, also the health-care efficiency measured. The fact that Greece and Portugal come up with relative good efficiency measures (see Section 4) might be due also to this factor. In so far as climate does play such a role, a correction of the (input or output) variables should be made in order to get a measure for the pure (man-made) health-care efficiency. Proxies for (good or bad) climate might be found in measures of environmental pollution. However, such a correction of the variables has not been done here.

Incomes in the medical sector

Incomes in the medical sector form a part of total health expenditure. If this component of costs differs between countries, it influences the resulting measure of efficiency. It is well known that, in fact, medical incomes differ substantially. For example, in the U.S. incomes of medical personnel are considerably higher than in other countries of comparable per capita income. But is the conclusion imperative to correct the total cost variable in order to make up for different levels of medical sector incomes?

Different levels of medical sector incomes – in terms of GDP per capita – may result from three reasons⁷. The first is different quality of the services between

countries. As far as this is the case, higher costs would translate into higher outputs. Thus, a correction would not be justified.

A second possible reason for different medical incomes is a different degree of monopoly on the side of the providers. Uncorrected figures of medical incomes would lead to lower efficiency – and that is what should be expressed by the scores in this case. Again, a correction would not be justified.

A third reason for differing medical incomes is a specific institutional setting. In the U.S., for example, it includes the system of self-financing of medical (university and non-university) studies as well as the high costs of insurance for liability of medical malpractice, which lead to relatively high incomes in the medical sector. Such institutional arrangements could, at least in principle, be changed in view of the example of other countries, so that lower medical incomes and lower health costs would occur. A lower measure of efficiency – before an institutional change or without such a change – would then be regarded as an appropriate characterisation of the health system. A correction of the input variable in order to make up for high costs was not indicated, similarly to the above-mentioned reasons of differences in quality and degree of monopoly.

With respect to the institutional setting, however, this position is not taken up here. It seems to be too far-fetched to assume that properties of the general national legal system or of the national educational system could be subject to change solely for lowering the costs of health care. Thus, corrections for differing income levels are reasonable – but only insofar as the differences result from different institutional settings, not insofar as they result from different qualities or different degrees of monopoly. Unfortunately, it does not seem possible to quantify and separate the three effects.

Assuming that different institutional settings which are not at the disposal of health-care authorities make up for the bulk of medical income differences, it seems to be reasonable to apply corrections of the total cost variable to make up for these cost differences.

The correction exercise should result in total costs of health care as if every country, and in each of the considered years, had the same level of medical incomes in terms of GDP per capita. A realistic value of the average income in the medical sector would be in the range of twice or three times the GDP per capita. But each arbitrary value will do. For the sake of simplicity, the correction is made as if average medical incomes in all countries, and always, were equal to GDP per capita.

In practical terms, the correction is made in the following way. (1) “total health employment” (OECD Health Data) is multiplied by GDP per capita. (2) From “total expenditure on health” (OECD Health Data) we subtract “total expenditure

⁷ I thank Peter Zweifel (see his Comment) for mentioning the first two of the reasons.

on medical services⁸ (OECD Health Data), then add the result of (1). What comes out is a measure for “total expenditure on health, as if average medical incomes were equal to GDP per capita” (in all countries and in any year). The corrected measure is, of course, then related to GDP. For the non-monetary variable health employment, there is no need for such a correction.

Age structure

The age structure of the population, differing considerably between countries, is of relevance for health-care costs because older people consume more medical services than younger people. It also influences the measure of health care efficiency, but it is not under the control of health-care authorities⁹. Thus, the input measure – the health expenditure variable as well as the health employment variable – must be corrected in order to make up for differences of the age structure. The method of correction is explained for the example of the health expenditure variable.

The correction exercise should result in such health expenditures as if every country had the same age structure in each year. In practical terms, the correction is made in the following way. For simplicity, we assume that every country has no young and no old, but only medium-aged people. The factually existing young and old people are converted to “medium-aged equivalents” with the help of their respective cost factors. These cost factors describe the annual average costs of health treatment of a young person relative to the costs of a middle-aged person, and the average costs of health treatment of an old person relative to, again, those of a middle-aged person. The country- and year-specific relation of the number of medium-aged equivalents to the factual total of the population is then used for the calculation of “total expenditure on health, corrected for differences in the age structure”.

Unfortunately, the cost factors mentioned probably differ between countries and change over time, but are not, to my knowledge, comprehensively known¹⁰. However, for some countries, estimates are published in the health economics literature¹¹. The best guess seems to be to use cost factors of 0.7 for young persons

⁸ “Total expenditure on medical services” is part of “Total expenditure on health” and can be regarded as labor costs of the health sector. The difference between the two variables could be considered as “capital costs” (not the expression of the OECD), although it also comprises items of current use and consumption.

⁹ In fact, health policy affects – and should affect – the age structure. The better the health policy fulfils its duties, the higher will become the share of old persons in the population – what is an extra-argument for correcting the input variable in order to make up for the cost effects of the age structure.

¹⁰ Unfortunately, the assumingly differing costs factors are of relevance in the context of this paper because they might be a result of reforms (or no reforms) of the provision of expensive medical treatments to the old-aged. See the Comment by Peter Zweifel.

¹¹ See, e.g., Hof, B. (2001), Fries, J. F. (1989), Kane, R.C. (1988), Verbrugge, L.M. (1984).

(0–14 years) relative to middle-aged persons, and of 3.0 for old persons (65+ years) relative to middle-aged persons. The cost factor for middle-aged persons is 1.0. All cost factors are (probably wrongly) assumed to be constant over time and to apply for all countries. The only alternative seems to be not to make any correction for differing age structures. I assume that the difference between the true and the assumed cost factors is smaller than the difference between the age structures. If this is true, doing no correction at all would result in a stronger distortion of the efficiency measure than a correction with identical cost factors. Thus, the correction will be made – in the simple way described.

Medical incomes and age structure

The input variable health employment has only to be corrected for the age structure. But the input variable total expenditure on health must be corrected for both medical incomes and age structure. The correction procedure corrects first for medical incomes and then, additionally, for the age structure.

5. The Results

All steps of the efficiency analysis use as output variable the reciprocal value of potential life years lost * 100,000 (OECD Health Data)¹². As inputs the following variables are used (each in a separate step of analysis):

1. Total expenditure on health, uncorrected
2. Total health employment per 1000 population, uncorrected (but implicitly corrected for medical incomes)
3. Total expenditure on health, corrected for medical incomes and age structure
4. Total health employment per 1000 population, corrected for age structure

Total expenditure on health is always as a percentage of GDP. Both variables are expressed in US\$ PPP¹³. For all steps of the analysis, the results relate to 1980, 1990, and 2000. The results for the health expenditure variable corrected only by one factor (only for medical incomes or only for age structure) are not presented.

At first (5.1) we ask whether the correction of the input variables makes any difference. In the following paragraphs (5.2 and 5.3) the uncorrected input variables and in paragraphs 5.4 and 5.5 the corrected input variables are used to evaluate the relative health-care efficiency of countries.

¹² The multiplicative factor of “*100,000” is added for a more convenient graphical presentation which, however, has not been used here.

¹³ It could have been expressed, however, in any currency, as long as it is the same for both costs and GDP.

The results of the FDH calculations are presented in Tables 1–4 in the Annex. Tables 5–8 in the Annex provide information about efficiency rankings for 1980, 1990 and 2000 and about the percentage change of efficiency from 1980–2000 (from 1990–2000 if earlier values are missing). The comparison of rankings, however, is only reasonably possible when the number of countries between years or between variables is equal. Thus, in several cases only a comparison of percentage changes of efficiency can be done.

1. Does the correction of the input variables make any difference?

We consider the Annex Tables 5 and 7 for health expenditures, uncorrected and corrected, respectively. What concerns the change of input efficiency, the difference is considerable for some countries. Switzerland fares much better with the corrected variables, while Australia, Finland and France have worse efficiency values. With respect to the change of output efficiency, Australia and United States fare better, while Finland has worse values.

To compare the health employment variable in the uncorrected and the corrected version, we consider Annex Tables 6 and 8. For the change of input efficiency, the correction exercise makes Portugal and United States fare slightly better, while Norway, Sweden, Switzerland and United Kingdom have worse values. With respect to the change of output efficiency, United Kingdom receives better values on the basis of the corrected variables, while Belgium, Italy and Netherlands fare worse.

Taken together, one can say that the differences between efficiency measures on the basis of uncorrected and corrected variables are quite substantial for several countries. Unfortunately, due to lack of data the correction exercise reduces the number of countries that can be compared. The following comparisons of efficiency are, thus, based on the uncorrected as well on the corrected variables.

2. Ranking of health-care efficiency in 2000, uncorrected input variables

We ask now for the ranking of health care efficiency in 2000 und consider the uncorrected input variables of health expenditure and health employment. The following Table 1 provides a summary of Annex Tables 5 and 6.

Table 1

Ranking of Health-Care Efficiency in 2000, Uncorrected Variables

Health expenditure, uncorrected		Health employment, uncorrected	
Input efficiency	Output efficiency	Input efficiency	Output efficiency
Dominant countries (first rank)			
Japan, Luxembourg, Mexico, Spain, Sweden, UK		Greece, Japan, Netherlands, Portugal, Spain, Sweden	
Countries in the lower five rankings			
USA, Portugal, New Zealand, Germany, Switzerland	Hungary, USA, Portugal, Slovak Rep., Poland	Finland, Norway, France, Switzerland, Germany	USA, Slovak. Rep., Finland, France, Germany

3. Changes of health-care efficiency 1980–2000, uncorrected input variables

We consider the changes of (relative) health-care efficiency between 1980 and 2000. Table 2 is again a summary of Annex Tables 5 and 6.

Table 2

Changes of Health-Care Efficiency 1980–2000, Uncorrected Variables

Health expenditure, uncorrected		Health employment, uncorrected	
Input efficiency	Output efficiency	Input efficiency	Output efficiency
Improvement by 20% and more			
Denmark, Sweden, Ireland, Hungary, Austria, Italy	Hungary, Ireland, Austria	Sweden, Australia, Canada, UK, USA, Netherlands, Switzerland, Italy	Portugal, Italy, Belgium
Deterioration by 10% and more			
Portugal, USA, Iceland, New Zealand, Switzerland	New Zealand, Iceland, Korea, Netherlands, Portugal, Poland, USA	Finland	USA

4. Ranking of health-care efficiency in 2000, corrected health employment variable

We consider now the ranking of countries according to the corrected input variable. A relevant number of countries is only reached with the corrected health employment variable. Thus, the corrected health expenditure variable is left aside. Table 3 is a summary of Annex Table 8.

Table 3

**Ranking of Health-Care Efficiency in 2000,
Corrected Health Employment Variable**

Health employment, corrected	
Input efficiency	Output efficiency
Dominant countries (first rank)	
Greece, Japan, Portugal, Spain, Sweden	
Countries in the lower five rankings	
Finland, Norway, France, Switzerland, UK	USA, Slovak Rep., Finland, France, Australia

5. Changes of health-care efficiency 1980–2000, corrected input variables

We ask now for the percentage change of efficiency between 1980 and 2000. The following Table 4 is a summary of Annex Table 8.

Table 4

**Changes of Health-Care Efficiency, 1980–2000,
Corrected Health Employment Variable**

Health employment, corrected	
Input efficiency	Output efficiency
Improvement by 20% and more	
Sweden, Australia, Canada, USA, Italy	Portugal, UK
Deterioration by 10% and more	
Finland, Ireland, France	USA

6. Summary

I now summarize Tables 1 to 4. They are based on three input measures, two uncorrected and one corrected input variable. All three measures are considered.

With respect to ranking Tables 1 and 3 show:

- There are three countries that rank at the top in all three measures: Japan, Spain and Sweden.
- Two countries rank at the top in two of the three measures: Greece and Portugal.

Lower ranks (no-top ranks) occur six times (for three input variables with input and output efficiency each). The five lowest ranks are considered. From six possible placements in one of the bottom five rankings:

- Finland and France: 4 times
- Germany, Slovak Republic, Switzerland, USA: 3 times
- Norway, Portugal: 2 times
- Australia, Hungary, Poland, New Zealand, UK: 1 time

With respect to change of efficiency, Tables 2 and 4 show that an improvement of efficiency (by 20% and more) and a deterioration of efficiency (by 10% and more) can occur, again, 6 times. From six possible placements in the category “improvement by 20% and more”:

- Italy, Sweden: 3 times
- Australia, Austria, Canada, Ireland, Hungary, Portugal, UK, USA: 2 times
- Belgium, Denmark, Netherlands: 1 time

From six possible placements in the category “deterioration by 10% and more”:

- USA: 3 times
- Finland, Portugal, Iceland, New Zealand: 2 times
- Ireland, France, Korea, Netherlands, Poland, Switzerland: 1 time.

The countries with the highest (relative) health-care efficiency in 2000 turn out to be Japan, Spain, Sweden, Greece and Portugal while lowest efficiency ranks are occupied by Finland, France, Germany, Slovak Republic, Switzerland and USA.

An improvement of efficiency between 1980 and 2000 happened specifically in Italy and Sweden while a deterioration occurred in USA.

7. Agenda for Further Research

The first question to be studied further must of course be whether the corrections that have been applied to the input variables really meet the point, how they can be refined and whether they should be changed. Of specific interest would be the possibility of a correction for the element of comfort as a joint product or unavoidable by-product of most health services. Moreover, it would be highly important to have a measure for total cost instead of total expenditure only, by correcting for loss of production due to absenteeism.

Secondly, in view of the results presented, it could be fruitful to search for the reasons why efficiency levels differ between countries and – even more important – why some of them have changed, in some cases considerably. It will be specifically interesting to know what type of policy measures – i.e., health-care reforms – has been the basis of improved efficiency. Some candidates of potentially successful reforms are immediately at hand, e.g. measures of rationing (either due to in-

creased co-payments¹⁴ or to longer waiting lists¹⁵ or to a more intensive use of Health Management Organisations).

Thirdly, only one output and one input have been used here. FDH analysis, however, can also be applied with more outputs and/or more inputs. But the question is, first, what is a reasonable kind of specification of the input-output relationship with more outputs and inputs than only one each, and second, is the database complete enough for other variables and for a larger number of countries than have been and could be analyzed in this paper.

Finally, as an alternative to the FDH method, another type of a non-parametric method, Data Envelopment Analysis (DEA), or a parametric estimation of a production function could be performed.

References

- Afonso, A./St. Aubyn, M. (2003): "Non-parametric Approaches to Public Education and Health Expenditure Efficiency in the OECD", paper presented during the annual conference of the International Institute of Public Finance, Prague.
- Comanor, W. S./Frech, H. E. III/Miller, R. D. (2003): "Is the United States an Outlier in Health Care and Health Outcomes?", paper presented at the CESifo Conference on Health and Economic Policy, June 2003, Munich.
- Culyer, A. J./Newhouse, J. P. (2001) eds., "Handbook of Health Economics", 2nd edition.
- Fries, J. F. (1989): "Erfolgreiches Altern. Medizinische und demographische Perspektiven", in: Erfolgreiches Altern. Bedingungen und Variationen, edited by M.M. Baltes, J. Kohli, Bern, Stuttgart, Toronto, 19–26.
- Gerdtham, U.-G./Jönsson, B. (2001): "International Comparisons of Health Expenditure", in Handbook of Health Economics, edited by A.J. Culyer and J.P. Newhouse.
- Gupta, S./Verhoeven, M. (2001): "The Efficiency of Government expenditure –Experiences from Africa", Journal of Policy Modelling 23, 433–467.
- Hof, B. (2001): "Auswirkungen und Konsequenzen der demographischen Entwicklung für die gesetzliche Kranken- und Pflegeversicherung", Cologne.
- Kane, R. C. (1988): "Beyond Caring: the Challenge to Geriatrics", Journal of the American Geriatrics Society 36.
- Newhouse, J. P. (1977): "Medical care expenditure: a cross country survey", Journal of Human Resources 12, 115–125.

¹⁴ For a theoretical assessment of co-payments and some relevant literature see, e.g., Osterkamp (2003, a); for an empirical application of co-payments to Germany see, e.g., Osterkamp (2003, b).

¹⁵ For an international comparison and an assessment of waiting lists see, e.g., Osterkamp (2002, b).

- Osterkamp, R.* (2003, a): “Public Health Insurance: Pareto-efficient allocative improvements through differentiated co-payment rates”, *European Journal for Health Economics* 4 (2), 79–84.
- (2003, b): “Hoehere Zuzahlungen und jeder profitiert – dargestellt für die deutsche GKV”, *List Forum für Wirtschafts- und Finanzpolitik* 29 (2), 159–178.
 - (2002, a): “Work Lost due to Illness – an International Comparison”, *CESifo Forum* 4, 36–40.
 - (2002, b): “Warten auf Operationen – ein internationaler Vergleich”, *ifo Schnelldienst* 55 (10), 14–21.
- Verbrugge, L. M.* (1984): “Longer Life but Worsening Health? Trends in Health and Mortality of Middle-Aged and Older Persons”, *Milbank Memorial Fund Quarterly* 62, 474–519.
- Zweifel, P.* (2004) Comment on “Is There a Hold-up of Health-Care Reforms in Europe?”, ARGE Conference Berlin, 2004

Annex 1: FDH Calculations*Table 1***Input Variable: Health Expenditure, Uncorrected, 1980 – 2000**

	1980		1990		2000	
	Input	Output	Input	Output	Input	Output
	Efficiency		Efficiency		Efficiency	
Australia	0.771	0.757	0.756	0.758	0.854	0.786
Austria	0.711	0.636	0.831	0.691	0.913	0.793
Belgium	0.844	0.759	0.797	0.703	0.874	0.794
Canada	0.761	0.740	0.656	0.751	0.826	0.823
Czech Republic			0.960	0.925	0.901	0.835
Denmark	0.593	0.777	0.694	0.672	0.904	0.805
Finland	0.844	0.871	0.756	0.656	0.955	0.976
France			0.686	0.702	0.785	0.733
Germany	0.621	0.702	0.694	0.703	0.708	0.768
Greece	0.818	0.796	0.797	0.793	0.777	0.735
Hungary			0.625	0.366	0.836	0.521
Iceland	1.000	1.000	0.738	0.827		
Ireland	0.643	0.758	0.787	0.715	0.875	0.946
Italy			0.738	0.785	0.927	0.941
Japan	1.000	1.000	1.000	1.000	1.000	1.000
Korea			1.000	1.000	0.949	0.856
Luxembourg	0.915	0.857	0.787	0.653	1.000	1.000
Mexico			1.000	1.000	1.000	1.000
Netherlands	0.853	0.938	0.738	0.844	0.884	0.818
New Zealand	0.915	0.927	0.855	0.636	0.700	0.750
Norway	0.928	0.915	0.766	0.807	0.987	0.866
Poland			0.906	0.741	0.903	0.662
Portugal	0.964	0.654	0.774	0.526	0.622	0.575
Slovak Republic				0.982	0.655	
Spain	1.000	1.000	0.881	0.729	1.000	1.000
Sweden	0.727	0.953	0.720	0.910	1.000	1.000
Switzerland	0.816	0.861	0.694	0.787	0.710	0.899
United Kingdom	0.964	0.919	0.983	0.746	1.000	1.000
United States	0.621	0.636	0.496	0.580	0.427	0.571

Table 2

Input Variable: Health Expenditure, Uncorrected, 1980 – 2000

	1980		1990		2000	
	Input	Output	Input	Output	Input	Output
	Efficiency		Efficiency		Efficiency	
Australia	0.321	0.757	0.427	0.758	0.810	0.827
Austria						
Belgium	0.559	0.692	0.659	0.886		
Canada			0.367	0.751	0.728	0.807
Czech Republic						
Denmark						
Finland	0.377	0.793	0.418	0.656	0.321	0.723
France			0.519	0.702	0.476	0.763
Germany			0.540	0.703	0.554	0.799
Greece	1.000	1.000	1.000	1.000	1.000	1.000
Hungary						
Iceland						
Ireland			0.711	0.902	0.670	0.911
Italy	0.596	0.757	0.763	0.990		
Japan	1.000	1.000	1.000	1.000	1.000	1.000
Korea						
Luxembourg						
Mexico						
Netherlands	0.667	0.938	0.968	0.844	1.000	1.000
New Zealand			0.794	0.801		
Norway			0.326	0.807	0.347	0.828
Poland						
Portugal	0.989	0.682	1.000	1.000	1.000	1.000
Slovak Republic					0.605	0.612
Spain					1.000	1.000
Sweden	0.285	0.953	0.408	0.910	1.000	1.000
Switzerland	0.360	0.861	0.265	0.787	0.516	0.882
United Kingdom	0.433	0.762	0.570	0.746	0.836	0.802
United States	0.388	0.636	0.431	0.580	0.619	0.571

Table 3

**Input Variable: Health Expenditure
Corrected for Medical Incomes and Age Structure, 1980 – 2000**

	1980		1990		2000	
	Input Efficiency	Output	Input Efficiency	Output	Input Efficiency	Output
Australia	0.858	0.807	0.798	0.899	0.874	0.956
Austria						
Belgium						
Canada			0.720	0.890	0.799	0.974
Czech Republic						
Denmark						
Finland	0.971	0.845	0.817	0.777	0.615	0.873
France			0.841	0.833	0.671	0.882
Germany					0.632	0.923
Greece						
Hungary						
Iceland						
Ireland						
Italy			1.000	1.000		
Japan						
Korea						
Luxembourg						
Mexico						
Netherlands	1.000	1.000	1.000	1.000	1.000	1.000
New Zealand						
Norway			0.515	0.956		
Poland						
Portugal						
Slovak Republic					0.852	0.612
Spain					1.000	1.000
Sweden						
Switzerland			0.611	0.933	1.000	1.000
United Kingdom					0.737	0.927
United States	0.902	0.677	0.694	0.688	0.577	0.689

Table 4

Input Variable: Health Employment, Corrected For Age Structure, 1980–2000

	1980		1990		2000	
	Input	Output	Input	Output	Input	Output
	Efficiency		Efficiency		Efficiency	
Australia	0.293	0.757	0.396	0.758	0.757	0.792
Austria						
Belgium	0.582	0.692	0.681	0.703		
Canada			0.343	0.751	0.686	0.807
Czech Republic						
Denmark						
Finland	0.370	0.793	0.414	0.656	0.297	0.723
France			0.520	0.702	0.454	0.763
Germany			0.564	0.703	0.545	0.799
Greece	1.000	1.000	1.000	1.000	1.000	1.000
Hungary						
Iceland						
Ireland			0.651	0.902	0.557	0.911
Italy	0.597	0.952	0.793	0.990		
Japan	1.000	1.000	1.000	1.000	1.000	1.000
Korea						
Luxembourg						
Mexico						
Netherlands	0.708	0.938	0.985	0.844	0.971	0.865
New Zealand			0.732	0.801		
Norway			0.362	0.807	0.349	0.828
Poland						
Portugal	0.937	0.682	1.000	1.000	1.000	1.000
Slovak Republic					0.526	0.612
Spain					1.000	1.000
Sweden	0.346	0.953	0.475	0.910	1.000	1.000
Switzerland	0.408	0.861	0.273	0.787	0.482	0.882
United Kingdom	0.456	0.762	0.599	0.746	0.501	0.927
United States	0.372	0.636	0.414	0.580	0.631	0.571

Annex 2: Rankings and Changes of Efficiency

Table 5

Input Variable: Health Expenditures, Uncorrected

	Input efficiency			Change of input	Output efficiency			Change of output
	1980 Ranking	1990 Ranking	2000 Ranking		1980 Ranking	1990 Ranking	2000 Ranking	
Australia	14	16	19	10.7%	16	12	19	3.7%
Austria	17	9	12	28.4%	20	21	18	24.7%
Belgium	10	10	18	3.5%	14	19	17	4.6%
Canada	15	26	21	8.6%	17	13	14	11.2%
Czech Republic		5	15	-6.1%		4	13	-9.7%
Denmark	21	22	13	52.3%	13	22	16	3.6%
Finland	10	16	9	13.2%	9	23	7	12.1%
France		25	22	14.4%		20	23	4.3%
Germany	19	22	25	14.0%	18	18	20	9.3%
Greece	12	10	23	-5.1%	12	9	22	-7.6%
Hungary		27	20	33.7%		28	28	42.2%
Iceland		1	18	-26.3%		1	7	-17.3%
Ireland	18	12	17	36.1%	15	17	8	24.8%
Italy		18	11	25.7%		11	9	19.8%
Japan	1	1	1		1	1	1	
Korea		1	10	-5.1%		1	12	-14.4%
Luxembourg	7	12	1	9.3%	11	24	1	16.6%
Mexico		1	1			1	1	
Netherlands	9	18	16	3.6%	5	6	15	-12.8%
New Zealand	7	8	26	-23.5%	6	25	21	-19.2%
Norway	6	15	7	6.4%	8	8	11	-5.4%
Poland		6	14	-0.3%		15	24	-10.7%
Portugal	4	14	27	-35.5%	19	27	26	-12.2%
Slovak Republic			8				25	
Spain	1	7	1	1	16	1		
Sweden	16	21	1	37.5%	4	5	1	4.9%
Switzerland	13	22	24	-12.9%	10	10	10	4.4%
United Kingdom	4	4	1	3.7%	7	14	1	8.8%
United States	19	28	28	-31.1%	21	26	27	-10.2%
Number of countries	21	28	28	25	21	28	28	25

Table 6

Input Variable: Health Employment, Uncorrected

	Input efficiency			Change of input efficiency	Output efficiency			Change of output efficiency
	1980 Ranking	1990 Ranking	2000 Ranking		1980 Ranking	1990 Ranking	2000 Ranking	
Australia	11	13	8	152.0%	8	12	10	9.2%
Belgium	6	8		17.8%	10	7		28.1%
Canada		16	9	98.5%		13	11	7.5%
Finland	9	14	17	-14.7%	6	17	15	-8.8%
France		11	15	-8.3%		16	14	8.7%
Germany		10	13	2.7%		15	13	13.6%
Greece	1	1	1		1	1	1	
Ireland		7	10	-5.8%		6	7	1.0%
Italy	5	6		28.0%	8	4		30.7%
Japan	1	1	1		1	1	1	
Netherlands	4	4	1	50.0%	4	8	1	6.6%
New Zealand		5				10		
Norway		17	16	6.4%		9	9	2.6%
Portugal	3	1	1	1.1%	11	1	1	46.7%
Slovak Republic			12				16	
Spain			1				1	
Sweden	12	15	1	250.7%	3	5	1	4.9%
Switzerland	10	18	14	43.4%	5	11	8	2.4%
United Kingdom	7	9	7	93.2%	7	14	12	5.3%
United States	8	12	11	59.5%	12	18	17	-10.1%
Number of countries	12	18	17	15	12	18	17	15

Table 7

**Input Variable: Health Expenditure,
Corrected for Medical Incomes and Age Structure**

	Input efficiency			Change of input	Output efficiency			Change of output
	1980 Ranking	1990 Ranking	2000 Ranking		1980 Ranking	1990 Ranking	2000 Ranking	
Australia	4	5	4	1.9%	3	5	5	18.5%
Canada		6	6	10.9%		6	4	9.5%
Finland	2	4	10	-36.7%	2	8	9	3.2%
France		3	8	-20.2%		7	8	5.9%
Germany			9				7	
Italy		1				1		
Netherlands	1	1	1		1	1	1	
Norway		9				3		
Slovak Republic			5				11	
Spain			1				1	
Switzerland		8	1	63.7%		4	1	7.2%
United Kingdom			7				6	
United States	3	7	11	-36.1%	4	9	10	1.8%
Number of countries	4	9	11	6	4	9	11	6

Table 8

Input Variable: Health Employment, Corrected for Age Structure

	Input efficiency			Change of input efficiency	Output efficiency			Change of output efficiency
	1980 Ranking	1990 Ranking	2000 Ranking		1980 Ranking	1990 Ranking	2000 Ranking	
Australia	12	15	7	158.1%	9	11	13	4.6%
Belgium	6	7		16.9%	10	15		1.6%
Canada		17	8	100.0%		12	11	7.5%
Finland	10	14	17	-19.7%	7	17	15	-8.8%
France		11	15	-12.8%		16	14	8.7%
Germany		10	11	-3.4%		14	12	13.6%
Greece	1	1	1		1	1	1	
Ireland		8	10	-14.5%		6	7	1.0%
Italy	5	5		32.8%	4	4		4.0%
Japan	1	1	1		1	1	1	
Netherlands	4	4	6	37.1%	5	7	9	-7.8%
New Zealand		6				9		
Norway		16	16	-3.7%		8	10	2.6%
Portugal	3	1	1	6.8%	11	1	1	46.7%
Slovak Republic			12				16	
Spain			1				1	
Sweden	11	12	1	189.2%	3	5	1	4.9%
Switzerland	8	18	14	18.0%	6	10	8	2.4%
United Kingdom	7	9	13	9.8%	8	13	6	21.7%
United States	9	13	9	69.9%	12	18	17	-10.1%
Number of countries	12	18	17	15	12	18	17	15

Annex 3: Raw Data*Table 9***Raw Data**

	Potential life years lost, all causes. younger than 70. per 100.000 popula- tion			Total expenditures on health, in % of GDP			Total health empl. per 1000 population (head count)		
	1980	1990	2000*	1980	1990	2000*	1980	1990	1999**
Australia	6104	4669	3771	7	7.8	8.9	28	31.6	32.1
Austria	7272	5125	3891	7.6	7.1	8			
Belgium	6684	5037		6.4	7.4	8.7	16.1	20.5	
Canada	6250	4716	7.1	9	9.2	36.8	35.7		
Czech Republic		6838	4776		5	7.1			
Denmark	5950	5270		9.1	8.5	8.3			
Finland	5828	5400	4085	6.4	7.8	6.7	23.9	32.3	46.7
France	6457	5041	4088		8.6	9.3		26	31.5
Germany	6583	5036	3907	8.7	8.5	10.6		25	29.4
Greece	5809	4463	4007	6.6	7.4	9.4	9	13.5	15
Hungary	9762	9664	7653			6.7			
Iceland	5075	4283		6.2	8	9.1			42.3
Ireland	6102	4950	4221	8.4	6.1	6.4		19	22.4
Italy	6104	4508	3415		8	8.2	15.1	17.7	
Japan	4623	3541	3120	6.4	5.9	7.6	14.4	21.5	28.4
Korea		6322	4657		4.8	5.9			
Luxembourg	6505	5419	3987	5.9	6.1	5.6			13
Mexico		9627			4.5	5.6			
Netherlands	4926	4197	3605	7.5	8	8.6	21.6	22.2	26
New Zealand	7015	5570		5.9	6.9	8		17	
Norway	5054	4388	3608	6.9	7.7	7.7		66	75
Poland	9272	8531	6026		5.3	6.2			
Portugal	8523	6736	5072	5.6	6.2	9	9.1	11	13.3
Slovak Republic			6086			5.7			22
Spain	5578	4857	3844	5.4	6.7	7.5			16.3
Sweden	4851	3892	2987	8.8	8.2	8.4	50.5	52.7	35.2
Switzerland	5367	4499	3386	7.6	8.5	10.7	40	51	55
Turkey				3.3	3.6				
United Kingdom	6067	4744	3889	5.6	6	7.3	20.8	23.7	31.1
United States	7274	6102	5229	8.7	11.9	13.1	23.2	31.3	36.2

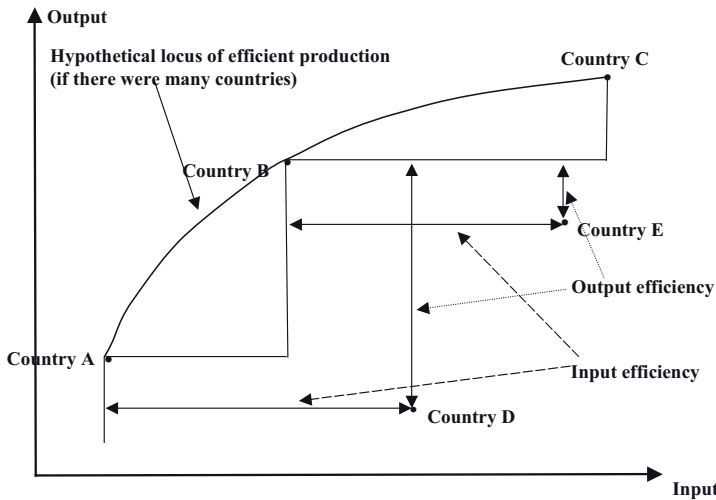
* Some information relates to 1999. – ** No information available for 2000, only for 1999.

Source: OECD Health Data, 2003.

Annex 4: How to Apply FDH Analysis

The procedure for applying FDH analysis is explained in two steps. At first, the simplest case of one input and one output only is explained with the help of the graph. It follows a description of the general case of multiple inputs and multiple outputs.

1. The Simplest Case



In the graph, the values for the input and output variables of countries A to E are empirically given. A, B and C are dominant countries, because there is no other country which uses less (or at least not more) from the input *and* gets more (or at least not less) from the output than countries A to C. Countries D and E are non-dominant countries because there is at least one country which dominates them. E is dominated by B, D is dominated by A and B.

To calculate the input and output efficiency scores for country E is straightforward. E is dominated in both respects by country B. Thus, the input efficiency score of country E, designated by $IES(E)$, is given by

$$IES(E) = I(B)/I(E)$$

The output efficiency score of country E, designated by $OES(E)$ is given by

$$OES(E) = O(E)/O(B)$$

$I(B)$, $I(E)$, $O(E)$ and $O(B)$ designate the inputs used and the outputs obtained, respectively, by both countries.

Country D is dominated by A and B. Thus, both input and output efficiency scores for D could be calculated in relation to either country A's input and output or country B's input and output. The FDH procedure is now to calculate the *input* efficiency score of country D by relating its input to the input of that country whose *output* is closest to (i.e. the least higher than) the output of country D. In the graph, this is country A. The *output* efficiency score of country D is calculated by relating its output to the output of that country whose *input* is closest to (i.e. the least less than) the input of country D. In the graph, this is country B. Thus:

$$IES(D) = I(A)/I(D)$$

$$OES(D) = O(D)/O(B)$$

2. The General Case

In the general, i.e. multi-input and multi-output, case we may have k inputs $x(x_1, \dots, x_k)$, m outputs $y(y_1, \dots, y_m)$ and n countries, whereby countries $n_1, \dots, n_d$ designate those countries which are dominant to country i whose efficiency scores are to be calculated. The input efficiency score of the non-dominant country i is calculated by:

$$IES(i) = \underset{n=n_1, \dots, n_d}{MIN} \quad \underset{j=1 \dots k}{MAX}(x_j(n)/x_j(i))$$

The output efficiency score of the non-dominant country i is calculated by:

$$OES(i) = \underset{n=n_1, \dots, n_d}{MIN} \quad \underset{j=1 \dots m}{MAX}(y_j(i)/y_j(n))$$

For a more elaborate presentation the reader may refer to Gupta and Verhoeven (2001) or to the Comment to this paper by Zweifel (2004).

Health-Care Efficiency in OECD Countries

Comment

By Peter Zweifel*

Attempts at health care reform are abundant throughout the developed world (see Busse and Schlette [2003] for an excellent survey). Yet, the suspicion lingers that these initiatives fall short of improving the efficiency of the healthcare systems concerned, resulting in a backlog of (effective) reforms. However, it is no simple task to substantiate such a suspicion. In his contribution, Rigmor Osterkamp tackles the challenge by applying a rather novel method – the Free Disposable Hull or FDH analysis – for establishing an efficient frontier, developed by Tulkens (1993). He tests whether European countries have tended to approach this frontier over the years 1980 to 2000, thus eliminating inefficiency in their healthcare sectors. He identifies Finland, France, and Germany as countries characterized by low initial efficiency scores while not improving them over time to conclude that there may be a hold-up of healthcare reforms in these three countries.

This comment focuses on three things. First, it provides a quick sketch of FDH analysis, comparing it to alternatives. Second, it addresses issues of specification of FDH analysis in the present context. Third, it raises a few questions about the results obtained.

FDH analysis as one method for establishing an efficient frontier

Establishing an efficient frontier for the case of one output only can be effected through estimating a production function. A pioneering attempt in the health domain was made by Auster et al. (1969), who related age-adjusted mortality rates in the United States to different types of healthcare inputs, income, education, alcohol and tobacco consumption, and several environmental factors. The alternative is to evoke duality theory, which states that assuming cost minimization by agents, the cost function contains the same information as the production function. In principle, a healthcare expenditure (HCE) function can thus be interpreted as a cost function; as such, it should only contain the output level achieved (a measure of health status in the present case) and factor prices (Mas-Colell et al. 1995, ch. 3.G

* Socioeconomic Institute of the University of Zurich, Hottingerstr. 10, CH-8032 Zurich.

and 5.C) as arguments. Although estimated HCE functions usually do not adhere to this theoretical prescription, they have been used for gauging the inefficiency of public HCE (Gerdtham et al. 1995).

As soon as there are multiple outputs, the scalar representation of the cost function is much easier to specify and estimate than a production correspondence. The additional outputs considered can simply be entered as additional explanatory variables, with flexible forms (e.g. the translog) providing guidance for specification. The downside of this easy generalization, especially in health care, is the poor quality of data on factor prices. Ambulatory care practices hardly provide data of sufficient detail for determining independent physician's implicit wage rates, while personnel of widely differing qualifications is often aggregated in the hospital sector. In addition, public hospitals traditionally do not have to account for their capital use and depreciation. Without this information, their user cost of capital cannot be determined.

In view of this difficulty, nonparametric methods for establishing efficient frontiers have become popular. A well-known method is Data Envelopment Analysis (DEA), which dates back to Charnes et al. (1978). It is based on the idea of associating efficiency with the maximization of a generalized distance between inputs and outputs. The efficiency frontier must be convex. In its input-oriented variant, a decision-making unit l (from an universe comprising L units) is tested for its efficiency by asking the question, "By how much can all of its inputs X_l be scaled down, provided it still uses the same linear combination of inputs and outputs as all the other units?". The underlying free disposal assumption means that reducing inputs does not engender any costs of disposal. Formally, the optimization problem (assuming constant returns to scale) is the following linear program (Banker et al., 1984):

$$\begin{aligned}
 & \min \theta_l, S, T. \\
 (1) \quad & X\lambda \leq \theta_l X_l \\
 & Y\lambda \geq Y_l \\
 & 0 \leq \lambda \leq 1.
 \end{aligned}$$

Here, θ_l is the scaling factor applied to unit l , i.e. its efficiency score, X and Y are the $(K \times L)$ input and $(M \times L)$ output matrices when the number of inputs is K and of outputs is M . The $(L \times 1)$ vector λ denotes the weights used for weighting the L units. Efficient units have an efficiency score of one ($\theta = 1$) because their inputs cannot be reduced any further. With an increasing number of inputs and outputs, more and more units define their own facet of the efficiency frontier, making them efficient for lack of comparison. The discriminatory power of DEA is reduced.

In formulation (1), DEA permits to split inefficiency into a scale component (units are either too small or too large to be efficient) and a technical component (their allocation of outputs and inputs is not efficient). Moreover, the dual of (1)

yields shadow prices of inputs and outputs. This implies that trade-offs are possible: A unit can attain a high efficiency score even though it is not productive in one dimension as long as it is productive in another, highly-valued one. Also, shadow prices can be constrained to realistic values in order to enhance the discriminatory power of DEA (Steinmann and Zweifel 2003). Alternatively, shadow prices can be compared to market prices to check the realism of the linear program specified.

The alternative used by Osterkamp, Free Disposal Hull (FDH analysis), also assumes free disposal of inputs. It proceeds in three steps. First, input and output matrices X and Y are scanned for combinations of values that make units dominant in relation to the unit l considered. These units must have a higher output-input ratio for all outputs m and inputs k , compared to unit l . Only units dominating l are retained because they may be part of the efficiency frontier. Second, for an input-oriented efficiency score, a matrix Z ($K \times L'$) is constructed that contains the input quantities of the L' dominating units relative to unit l , i.e. $Z_{k,l'} = X_{k,l'} / X_{k,l}$. For a given dominating unit l' , this yields a $(K \times 1)$ vector of ratios which are necessarily equal to or less than one. The question arises which element to select in order to characterize the position of unit l relative to the “challenger” l' . The maximum, i.e. the input ratio most favorable to l is retained. This is consistent with optimizing behavior in the sense that unit l strives to be as close to the efficient frontier as possible at least with regard to one input. Step three consists in constructing the $(1 \times L')$ vector containing those maximum values. Again, the question arises of which element to choose in order to characterize the efficiency of unit l relative to the efficiency frontier. Clearly the minimum of these values must be retained because among the L' “challengers”, the one with the minimum relative input quantity defines the frontier.

In all, these three steps result in the input-oriented efficiency score

$$(2) \quad i_l = \min_{l' \in L'} \max_{k \in K} \left[Z_{k,l'} = \frac{X_{k,l'}}{X_{k,l}} \right]$$

For obtaining an output-oriented efficiency score, a $(M \times L')$ matrix W is constructed that contains all the output ratios of the “challengers” relative to unit l . Assuming optimizing behavior by l , the best ratio is the minimum one this time. Among those L' units that may qualify to be efficient, however, the one featuring maximum value of relative output is the appropriate one. Therefore, the output-oriented efficiency score is given by

$$(3) \quad o_l = \max_{l' \in L'} \min_{m \in M} \left[W_{m,l'} = \frac{Y_{m,l'}}{Y_{m,l}} \right]$$

Since no linear combination of units is involved in the construction of the FDH scores, there is no need to impose convexity (contrary to DEA). On the other hand, the argument leading to eq. (2) shows that trade-offs are ruled out. FDH does not

yield shadow prices that may be used to increase discriminatory power of the analysis or to check the realism of the specification.

Specification of FDH in the paper

Since discriminatory power cannot be increased through restrictions on shadow prices, parsimony is of particular importance in FDH analysis. Yet when the number of outputs is cut down to one (the inverse of life years lost due to premature mortality in this case), an important advantage of FDH analysis over the production function approach is lost. Estimation of a production function would generate evidence on returns to scale and possibly elasticities of substitution. The latter concept is of interest because it shows how different inputs could be substituted for attaining a given health outcome. By their very construction, both FDH efficiency scores must remain silent on this issue.

With only 20 European and 7 non-European countries and missing data in each year studied, parsimony may indeed dictate the use of only one input variable as well. The author alternatively uses HCE and employment in health care. From a strictly production-theoretic viewpoint, employment would be preferable; however, the author justifiably argues that HCE aggregates over several inputs of differing quality. But then he goes on to correct for differences in medical incomes, constructing a quantity “total expenditure on health as though average medical incomes were equal to GDP per capita”. This correction is questionable on several counts. First, the variable above should be “... as though average medical incomes were equal to GDP per individual in the labor force” since differences of participation rates bias the comparison. Second, these income differences might be due to quality differences; correcting the data for them therefore militates against the justification just given for using HCE data. Third, HCE may be high for a given health outcome precisely because medical incomes (or rather fees) are excessive. One important element of reform is to make markets for healthcare services (and medical services in particular) more competitive. Indeed, there is some comparative evidence suggesting an effect of cartelization on medical incomes (Zweifel and Grandchamp 2002). This downward adjustment of HCE in an expensive country thus causes measured efficiency to be high in a country that needs to open up healthcare markets to more competition – an important reform indeed.

The HCE variable is then subjected to a second correction, based on estimates of average cost factors. E.g., HCE of persons aged 65+ is set to the threefold of HCE of the population 15–64. Apart from the first correction, the input quantity then becomes “... HCE as though both the age structure and the effect of age on cost were identical across countries”. However, to the extent that the (very) old are treated using expensive medical technology, a reform might aim at changing incentives of both patients and healthcare providers as to reduce this effect of age. This second adjustment thus again serves to undermine the usefulness of the efficiency

scores as indicators of a reform backlog. The alternative would be to adjust the output quantity for differences in the age structure. While this presumably would change little in the present context with only one input, it does make a difference as soon as there are inputs besides HCE, e.g. education. Efficiency could then be reflected in the way spending on both education and health care contributes to a reduction of age-adjusted life years lost.

A final specification step consists in replacing HCE by the health share HCE / GDP. While this has the considerable advantage of obviating adjustments for international differences in purchasing power, it does imply that a one percent increase in HCE associated with a one percent increase in income is not expected to have an effect on population health. Thus, increased income would have to be associated with lower health status *ceteris paribus* – a finding by Auster et al. (1969) for the United States which however has not been corroborated by later studies. Moreover, the output elasticities with regard to HCE and GDP would have to be equal in absolute value, which constitutes a very restrictive assumption.

Results

In all, HCE is subjected to a total of three adjustments in an attempt to restrict the number of inputs in this FDH analysis to one, the motivation being that with more inputs, discriminatory power presumably would be lost. However, these adjustments also reduce discriminatory power. As evidenced in table 1 of the paper, just 2 out of 13 countries for which input and output efficiency scores could be calculated are fully efficient ($i_l = 1$, $o_l = 1$) when untransformed HCE serves as the input variable. With transformed HCE, this number rises to 4 out of 13.

Some changes over time are also baffling. For one, i_l for Denmark is at a low 0.593 in 1980, rises slightly to 0.694 in 1990, and jumps to 0.904 in 2000 (Table 1 of the Annex, HCE not corrected for medical income and age). A similar pattern obtains for Sweden, with 0.727, 0.720, and 1.000. According to the maintained hypothesis of the paper, Denmark and Sweden would have had to institute major reforms during the 1990s. They did strengthen the role of regional health agencies as purchasers of healthcare services, but it is doubtful that this change should be the cause of such a marked increase in efficiency. For comparison, the United Kingdom during the 1990s had introduced a similar concept by creating fund-holding practices as purchasers of specialist and hospital services. Yet its score merely increases from 0.983 to 1.000 during that decade.

When HCE is corrected for medical income and age, the Netherlands is fully efficient in all three years ($i_l = 1$, $o_l = 1$). However, pursuant the Dekker report, the Dutch government sought to substantially increase the share of the population covered by private insurance. One would expect this reform to cause a change in the efficiency scores. Finally, i_l for Switzerland jumps from 0.611 in 1990 to 1.000 in

2000. Although a new federal law (effective 1996) injected more competition into health insurance, insurers still lack the freedom to contract with either physicians or hospitals. This marked efficiency improvement thus seems to lack justification. Altogether, these observations serve to shed some doubt on the validity of the FDH efficiency scores.

Concluding remarks

The contribution by Rigmar Osterkamp addresses an important issue, viz. the measurement of the overall efficiency of healthcare systems. Given that data on factor prices are error-ridden or not even available in health care, nonparametric methods have an advantage. However, in this particular application with just one output, the comparative advantage of FDH analysis over the parametric alternative is not clear. A production function approach based on a merged sample but using time dummies would have made it possible us to relate health to several rather well-established determinants. Negative outliers that are robust with regard to these time dummies could then have been associated with countries that are characterized by a marked need for healthcare reform. Finally, the comparative advantage of FDH over DEA would have to be established as well. Using DEA, the need for reform could be indicated by a coincidence of scale inefficiency, allocative inefficiency, and deviations of actual relative prices (where observable) from relative shadow prices. This said, it is fair to point out that this contribution is the first to apply FDH analysis to the patchy data on the healthcare systems of industrial countries and to derive indicators of the likely backlog of reforms in several European countries.

References

- Auster / Leveson, I. / Sarachek, D. (1969): "The production of health, an exploratory study", *Journal of Human Resources* 4(4), 411–436.
- Banker, R. D. / Charnes, A. / Cooper, W. W. (1984): "Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis", *Management Science* 30, 1078–1092.
- Busse, R. / Schlette, S. eds. (2003), "Health Policy Developments", Bertelsmann Foundation Publishers, Gütersloh.
- Charnes, A. / Cooper, W. W. / Rhodes, E. (1978): "Measuring the efficiency of decision making units", *European Journal of Operations Research* 2, 429–444.
- Gerdtham, U. / Jönsson, B. / MacFarlan, M / Oxley, H. (1995): "New directions in health care policy", *OECD Health Policy Studies* 7, Paris.
- Mas-Collel, A. / Whinston, M. D. / Green, J. R. (1995): "Microeconomic Theory", Oxford University Press, New York.

- Steinmann, L. / Zweifel, P.* (2003): “On the (in)efficiency of Swiss hospitals”, *Applied Economics* 35, 361–370.
- Tulkens, H.* (1993): “On FDH analysis: some methodological issues and applications to retail banking, courts, and urban transit”, *Journal of Productivity Analysis* 4, 183–210.
- Zweifel, P. / Grandchamp, C.* (2002): “Measuring the effect of cartelization in medicine: an international study”, in *Measuring Market Power*, edited by D. J. Slottje, Ch. 3, North Holland-Elsevier, Amsterdam.

Participants

- | | |
|---|---|
| 1. Dr. <i>Dagmar Balve-Hauff</i> | BMW Berlin |
| 2. Prof. Dr. <i>Norbert Berthold</i> | Universität Würzburg |
| 3. Prof. Dr. <i>Charles Blankart</i> | Humboldt Universität, Berlin |
| 4. <i>Deborah Anne Bowen</i> | AEQ, Berlin |
| 5. Dr. <i>Rainer Brandl</i> | IZA, Bonn |
| 6. Dr. <i>Albert Caspers</i> | BMW Berlin |
| 7. Dr. <i>Berend Diekmann</i> | BMW Berlin |
| 8. Prof. <i>Karl Heinz Domdey</i> | IWVWW, Berlin |
| 9. <i>Henny Dumröse</i> | BMW Berlin |
| 10. Dr. <i>Werner Eichhorst</i> | IAB, Nürnberg |
| 11. Dr. <i>Georg Erber</i> | DIW Berlin |
| 12. Prof. Dr. <i>Mathias Erlei</i> | Universität Clausthal |
| 13. <i>Ulrike Fischer</i> | DIW Berlin |
| 14. Dr. <i>Wolf-Dieter Fuechsel</i> | BMW Berlin |
| 15. Prof. Dr. <i>Daniel Gros</i> | CEPS, Brüssel |
| 16. <i>Frank Grohmann</i> | BMF Berlin |
| 17. Dr. <i>Peter Haug</i> | Institut für Wirtschaftsforschung Halle |
| 18. <i>Reinhard Heck</i> | BMW Bonn |
| 19. Dr. Ph.D. <i>Carsten Hefeker</i> | HWHA Hamburg |
| 20. Dr. <i>Friedrich Heinemann</i> | ZEW Mannheim |
| 21. <i>Yvonne Heiniger</i> | HWHA Hamburg |
| 22. <i>Christian Heydecker</i> | BMW |
| 23. <i>Jochen Homann</i> | BMW Bonn |
| 24. Dr. <i>Gunter Kayser</i> | Institut für Mittelstandsforschung Bonn |
| 25. Dr. <i>Jochen Kluve</i> | RWI, Essen |
| 26. Dr. <i>Wilhelm Koll</i> | BMW Berlin |
| 27. Dr. <i>Rolf Kroker</i> | Institut der deutschen Wirtschaft, Köln |
| 28. Prof. Dr. <i>Rolf J. Langhammer</i> | Institut für Weltwirtschaft, Kiel |
| 29. Dr. <i>Konrad Lammers</i> | HWHA Hamburg |

- | | |
|--|--|
| 30. Dr. <i>Harald Legler</i> | Niedersächsisches Institut für Wirtschaftsforschung,
Hannover |
| 31. <i>Klaus-Peter Leier</i> | BMW Berlin |
| 32. <i>Marion Lemgau</i> | BMW Berlin |
| 33. <i>Elzbieta Linke</i> | HWHA Hamburg |
| 34. <i>Annett Löser</i> | BMW Berlin |
| 35. <i>Hendrik Luchtmeier</i> | BMW Berlin |
| 36. <i>Andreas Lux</i> | BMF Berlin |
| 37. Prof. Dr. <i>Helmut Matthes</i> | IWVWW Berlin |
| 38. <i>Ralf Messer</i> | ARGE, DIW Berlin |
| 39. <i>Martin Meurers</i> | BMW Berlin |
| 40. <i>Hans-Michael Meyer-Sebastian</i> | BMF Berlin |
| 41. Dr. <i>Gesa Mieke-Nordmeyer</i> | BMW Berlin |
| 42. <i>Andreas Mihm</i> | Frankfurter Allgemeine Zeitung |
| 43. <i>Martin Mindermann</i> | BMW Berlin |
| 44. <i>Klaus Müller</i> | BMW Berlin |
| 45. Dr. <i>Rigmar Osterkamp</i> | ifo, München |
| 46. Prof. Dr. <i>Rüdiger Pohl</i> | Institut für Wirtschaftsforschung Halle |
| 47. <i>Christian Pröbkuß</i> | BMW Berlin |
| 48. Dr. <i>Maike Richter</i> | BMW Berlin |
| 49. Dr. <i>Steffen Roth</i> | Institut für Wirtschaftspolitik, Köln |
| 50. <i>Aida Sayed-Ahmed</i> | DIW Berlin / Uni Tübingen |
| 51. Prof. Dr. <i>Ronnie Schöb</i> | Universität Magdeburg |
| 52. Dr. <i>Camilla Schloenbach</i> | BMF Berlin |
| 53. <i>Hildegard Stahmer</i> | HWHA Hamburg |
| 54. Prof. Dr. <i>Viktor Steiner</i> | DIW Berlin |
| 55. Prof. Dr. <i>Thomas Straubhaar</i> | HWHA Hamburg |
| 56. <i>Thies Thormählen</i> | BMF Berlin |
| 57. <i>Peer-Oliver Villwock</i> | BMW Berlin |
| 58. Dr. <i>Horst Würzburg</i> | BMW Berlin |
| 59. Prof. Dr. <i>Klaus F. Zimmermann</i> | DIW Berlin |
| 60. <i>Carsten Zinkau</i> | BMF Berlin |
| 61. Prof. Dr. <i>Peter Zweifel</i> | Universität Zürich |