CESifo 🎲 DICE REPORT
Journal for Institutional Comparisons

CESifo Group Munich

Home | Research | **Economic Facts and Figures** | Policy Debate | Publication Series | Conferences and Events | Media Relations | About Us

Survey Results | Forecasts | **Institutional Comparisons** | Ifo Economic Data | Educational Material | Survey Participation

Business and Financial Markets

Education and Innovation

Energy and Natural Environment

Infrastructure

Labour Market and Migration

Other Topics

Public Sector

Social Policy

Values

Ifo's Database for Institutional Comparisons in Europe (DICE)

Introduction

The database DICE is an additional service product of the Ifo Institute. DICE provides systematic information on economic institutions and regulatory systems as well as their economic effects. The information is presented as charts, tables and reports. The user can choose between current comparisons, archived documents from previous years and time series that show developments over time. In most cases, the 27 EU countries as well as some major non-EU countries are covered.

Why DICE?

More than anything else it is the institutional framework of an economy and the implied incentive structure that explain the welfare of a nation. In a time of rapidly expanding globalisation, where people, businesses, capital and goods are becoming increasingly mobile internationally, countries are compelled to compete on the basis of the institutions they have established. No country can afford not to compare its institutions with those of its neighbours, and all countries must aim for international benchmarks and best practices. International institutional data that allow a country to assess its own situation and to prepare reforms are in great demand.

Formats of DICE

The information which DICE contains will be presented in three different formats.

- DICE charts depict selected features of important new institutional-economic developments and are therefore of interest to a broad public.
- DICE tables cover a broad range of topics and offer more detailed and in-depth information.
- DICE reports highlight topical economic developments with brief texts built around graphical illustrations and tables. They are regularly published in the quarterly CESifo DICE Report.

CESifo DICE

Introduction
List of all institutional fields
Search form
New in DICE

Database links

Current folders
Archives

DICE • REPORT

Reports are regularly published in the quarterly DICE Report

Home | | Contact | Impressum | Privacy Policy | Sitemap | Search | Print view   © 1999-2009 Ifo Institute for Economic Research, Munich

DICE Database: www.cesifo-group.de/DICE

DICE Report

CULTURE IN ECONOMICS

WOMEN, CULTURE AND ECONOMICS

RAQUEL FERNÁNDEZ*

Over the last two hundred years there have been immense changes both in how societies view women and what women do relative to men. How and why this has happened has been the subject of a vast literature which I will not even begin to attempt to review here. Instead, I would like to address, at least in part, the question of whether changes in culture – by which I mean changes in a set of social beliefs and understandings – played an important role in this transformation.

Both economists and non-economists alike may be tempted to dismiss this question or stop reading this article at this point. The latter may consider the answer as evidently "yes" whereas the former may, in general, consider it a question that – like the existence of God – economics is not designed to answer. Yet, as hopefully this issue of DICE REPORT makes clear, some economists have lately made headway in studying the role of culture.[1] Culture, we would argue, and its role in questions relating to growth, technological progress, trade, crime, etc. – i.e., in contributing to any of the phenomena that economists find of interest – can and should be studied with the entire array of rigorous techniques and methodology available to economists.

Does culture play an independent role in traditional economic issues? Let me create a straw-man (we can call him homo materialisticus) as a way to clarify the way in which I would like to think about this question. Homo materialisticus believes that the great economic changes that the world has seen are basically a result of technological or other material change (e.g., the discovery of new lands). Culture – societal beliefs – exists and changes over time and space, but it is basically some flotsam of the mind with no independent role other than reflecting the underlying material reality. So, yes, attitudes towards women have changed, but this has been a reflection of technological change that altered what women do and thus how we think of them. It is this mindset that the work I discuss below seeks to challenge by providing evidence to disprove it.

It is not my intention here to propose and develop a more sophisticated model in which culture, institutions, history, technology and the general material environment interact to determine outcomes and how they all change over time. Nor will I attempt to answer the grand question of the role occupied by culture in transforming women's place in the economy. Instead, I will review some of the work that I and coauthors have done to show that cultural variations that exist over space matter to economic outcomes, particularly women's market work. If differences in culture matter, then it is not too much of a leap from there to thinking that changes in culture must matter as well.

How might culture affect women's participation in the formal labor market (working for a wage)? Note that this is a very different question than asking about how culture affects whether women work – historically, women have always worked. Restricting ourselves to the last one hundred or so years, over this time a dramatic change occurred in the types of work that women do relative to men and, particularly for married women, in their ability and inclination to work for a wage outside the home. It is illustrative to use numbers from the US to make this point as they span a long time period: the labor force participation in market work of white, married women between the ages of 25 and 44 who were not in agriculture increased from about 3 percent in 1880 to over 73 percent in 2000.[2]

---

* Professor of economics, New York University, Department of Economics.
[1] For a review of the literature see Fernández (2007) and Guiso et al. (2006).

[2] Numbers are from author's calculations in Fernández (2008). There have also been important changes in the work habits of single women, but in terms of labor force participation and hours worked (the easiest things to measure), it is the change in married women's labor supply that is the most dramatic.

At any moment in time, whether a (married) woman worked outside the home depended upon other things, on social (and individual) beliefs. These beliefs ranged from those concerning a woman's productivity, or regarding her emotional stability or intellectual ability or ambition, as well as concerns about her welfare, her marriage's stability or her children's welfare if she worked. These beliefs – these social anxieties – along with the associated social rewards and punishments from working have all been part of the social context in which women made their work decisions. Did this social context play a real role in influencing women's work decisions, however, or were these decisions simply based on the economic environment with the social narrative above playing an insignificant role? Below I will attempt to answer a modest version of this question by reviewing some recent work that shows that culture appears to play an important role in determining how much women work outside the home.

Taking what I have called an "epidemiological" approach to the analysis of culture, my research studies the descendants of immigrants from various countries of ancestry but who live in the same country to separate the role of culture from that of the technological and institutional environment. The reasoning underlying this approach is similar to that employed by epidemiologists. Much as epidemiologists might attempt to distinguish between the genetic versus environmental contributions to, for example, differences in the rate of heart disease between Japanese and American men by studying Japanese immigrants and their descendants in the US and comparing them to others living in the US, one can use a similar approach to distinguish between the impact of beliefs and the environment in women's work behavior. In the case of heart disease, finding convergence in the incidence of the disease between Japanese descendants and others would tend to support the theory that an environmental agent is responsible for the health differences between Japan and the US.[3] Similarly, if immigrants from a diverse range of countries and their descendants in the US display the same economic behavior, this would tend to lend support to a more "materialist" view of the world.
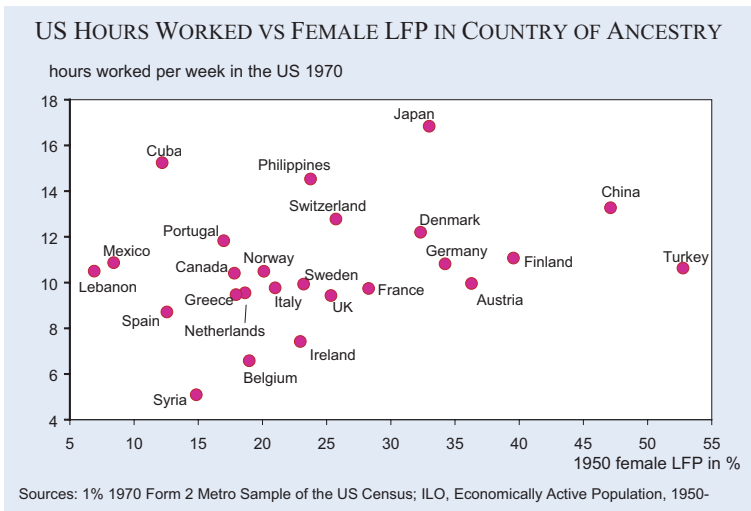
The use of an epidemiological approach is considerably more complex in economics than in medicine. There are many reasons why cultural differences may fail to play an important role in the life of immigrants or their descendants and yet be a very important determinant of behavior. First and foremost, culture is socially constructed and maintained. Even if Japanese cultural beliefs about how women should behave may have a major influence on women's work habits in Japan, the impact of these beliefs on Japanese women in the US may be significantly weaker since they occur in a social setting in which the social rewards and sanctions are different.[4] Second, immigrants are not randomly selected individuals from their country of ancestry and thus may not possess the "average" social beliefs. Third, although we study second-generation individuals in order to minimize the importance of differences in language ability and other immigrant "shocks", this strategy will be weakened by immigrant assimilation and thus tends to bias the findings against culture. Thus, altogether this strategy is biased against finding that culture plays a role in economic outcomes and hence can only be convincingly used to provide evidence that culture matters rather than to disprove it.

In Fernández and Fogli (2006; 2009) and Fernández (2007), we study the work and fertility outcomes of second-generation American women in 1970. Using the 1970 US Census, we construct a sample of married women (as this is the group of women for which one finds interesting differences in labor supply) between the ages of thirty and forty, and distinguished by the characteristic of having parents that were born outside the US. This results in a sample of over 6,750 married women from 25 countries of ancestry around the world.

An important challenge is how to measure culture. In Fernández and Fogli (2009) we use the female labor force participation (LFP) rate in 1950 in the father's home country. This variable reflects the beliefs that women held in 1950, in the father's home country, regarding the desirability of working, as well as employers' views towards employing women. It can also reflect social attitudes that may facilitate or hinder working (e.g., by affecting the availability of high-quality child care), as well as the degree of support within the family for a working wife, and of

---

[3] Interestingly, for an epidemiologist the environment would include culture, e.g., the greater propensity to eat fish, whereas economists are interested in distinguishing between culture (presumably relatively easily portable to a new economic environment) and the features of the environment that are less portable, e.g., institutions, market conditions, etc.

[4] Some groups manage to preserve a very different set of values from those of society on average despite the fact that they directly affect individual economic outcomes by effectively isolating themselves, e.g., the Amish or Hasidic Jews.

Figure 1

US HOURS WORKED VS FEMALE LFP IN COUNTRY OF ANCESTRY

hours worked per week in the US 1970

Sources: 1% 1970 Form 2 Metro Sample of the US Census; ILO, Economically Active Population, 1950-

second-generation American women work in the US other than through beliefs.

Figure 1 reproduces the raw correlation found in Fernández and Fogli (2009) between the average number of hours worked per week by women of different countries of ancestry in 1970 in the US and the 1950 female LFP in these countries. As can be seen, there is a positive correlation between these two variables. Before turning to the question of whether this relationship is causal, I will review additional evidence on the link between culture and women's work.

course local market and production conditions that determine wages as well as women's investment in human capital that determine their productivity. Thus the aggregate participation of women in the work force reflects culture, institutions and the economic environment.

The critical argument for using female LFP in 1950 as a proxy for culture is that, in the context of explaining the behavior of second-generation American in the US, it will not be contaminated by the contribution of institutions and the economic environment that influence its level in the country of ancestry. The asymmetry between culture and the other contributors rests on their degree of "portability." That is, whereas the cultural component of culture may be transmitted from parents (and perhaps the wider immigrant community) to their daughters (and reinforced perhaps by the immigrant community) and thus affects the beliefs of second-generation American women in 1970, the local material and institutional conditions of the country of ancestry in 1950 should not be transferable. Thus, although we don't formally use the female LFP variable as an instrument, in order for it to be a valid proxy for culture in our context, it needs to fulfill the same conditions. In particular, it should not have any direct effect on how much these

In Fernández (2007), I employ the epidemiological approach to study the relationship between culture and work, using social attitudes instead. I used the answers to questions posed in the World Value Survey (WVS) in 1990–91 that reveal attitudes towards women's work. In particular, individuals were asked to answer whether they strongly agreed, agreed, disagreed or strongly disagreed with the following statements:

1. Being a housewife is just as fulfilling as working for pay;
2. Having a job is the best way for a woman to be an independent person.

There is a great deal of variation across the European countries in how individuals reacted to these statements.[5] Figure 2 shows the percentage that strongly

Figure 2

BEING A HOUSEWIFE IS AS FULFILLING AS WORKING FOR PAY

% disagree or strongly disagree (WVS 1990)

Source: World Value Survey, 1990.
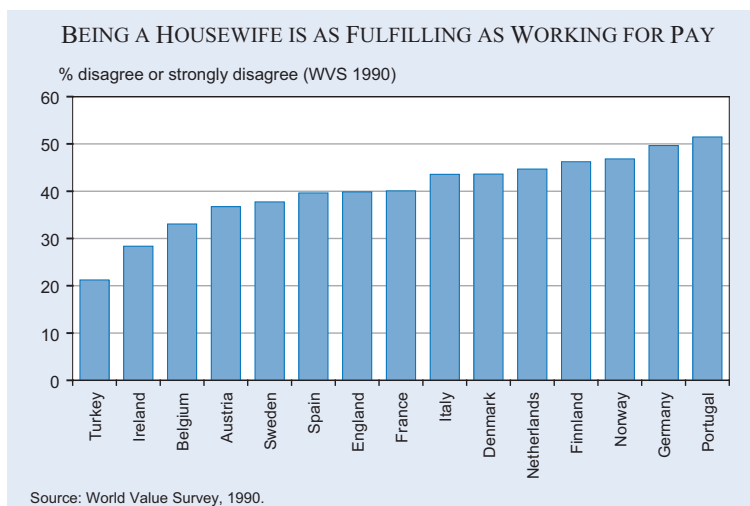
[5] The sample was restricted to European countries as the questions were not posed to the other countries in our US Census sample (with the exception of two others).

Figure 3

BEING A HOUSEWIFE IS AS FULFILLING AS WORKING FOR PAY

average hours worked per week in US in 1970

Sources: World Value Survey, 1990; 1% 1970 Form 2 Metro Sample of the US Census.
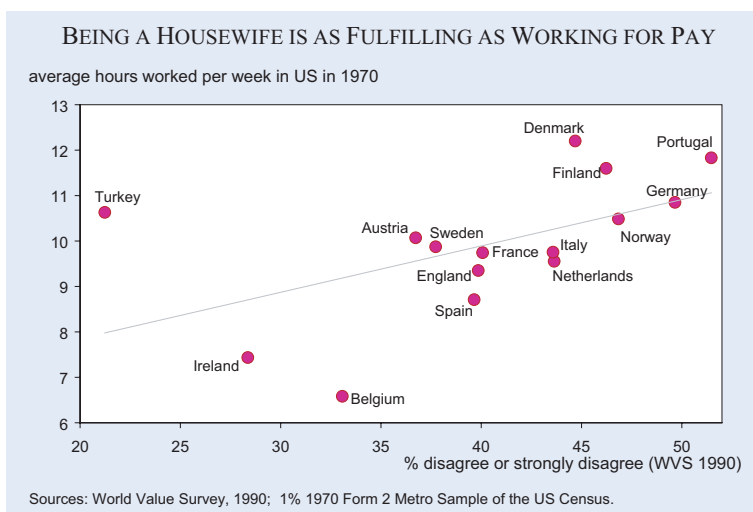
disagreed or disagreed with statement 1 above by country. The number ranged from a bit over 20 percent in Turkey to slightly over 50 percent in Portugal.

Rather than study whether the attitudes expressed by individuals across countries help explain the cross-country variation in women's work in Europe, I used the epidemiological approach and investigated instead whether the cross-country variation in attitudes was capable of explaining variation in the work behavior of the second-generation American women (with European ancestry) in 1970. Note that, once again, this strategy relied on using country-of-ancestry variables to explain why women work different amounts in *the same country* (the US) – hence eliminating other factors that could influence cross-country variation in women's labor market outcomes. For this to be a valid proxy for culture, we need for the attitudes expressed by individuals across European countries in 1990 should not have a direct influence on the work habits of second-generation American women in the US in 1970 other than through the common beliefs that may underlie both variables. Figure 3 shows the raw correlation between the average number of hours worked per week in 1970 in the US by women of different countries of ancestry and the percentage of individuals who in that same country in 1990 disagreed with statement 1.[6]

I next turn to the main challenge faced by this research strategy – the concern that an omitted variable is driving the positive correlations shown in Figures 1 and 3. As a fundamental first step, all the

regression analyses included a set of characteristics of married women that might vary in a systematic fashion across groups of ancestry. In particular, we controlled for the woman's age and education as well as the age and education of her husband and her husband's total income. Furthermore, we included fine geographic controls (over 100 standard metropolitan area dummies) to capture potentially systematic differences in the residence pattern of different ethnic groups might reside which could then be correlated with different market conditions. The correlation remained positive and statistically and economically significant.[7]

Despite the inclusion of individual-level controls, it remained possible that an omitted variable could be responsible for the results. If this variable is not culture than it needed to directly affect either the material cost or the benefit from working. The main suspect would be, therefore, unobserved human capital. That is, although we were controlling for a woman's education, some aspects of human capital might not be captured by this variable.

In Fernández and Fogli (2009) we investigated in depth the challenge posed by unobserved human capital. One possible avenue was that parental education differed systematically by country-of-ancestry and that work behavior of daughters reflected the transmission of this by channels not captured by formal education. As the US Census did not have information on parental education, we constructed proxies for parental human capital by obtaining the education level of the parent's immigrant group from the 1940 Census. We also used the General Social Survey, which contained information on parental education, to obtain an alternative (unfortunately, significantly smaller) sample of second-generation women. Furthermore, to attempt to capture quality as well as quantity measures of parental human cap-

---

[6] The results for statement 2 are very similar.

[7] In particular, after including all individual and geographic controls, a one-standard-deviation increase in female LFP in 1950 is associated with almost an 8 percent increase in hours worked in US in 1970. Similarly, a one-standard-deviation increase in the country marginal effects (obtained by regressing the individual response to statement 1 – coded agree = 1 and disagree = 0 – on a series of demographic variables and a country dummy for the 15 countries in our WVS European sample), is associated with a 6 percent decrease in average hours worked in the US in 1970.

ital, we included Hanushek and Kimko's (2000) international test-based measures of quality differentials in education across countries. The relationship between the cultural proxy and hours worked survived all these tests. Lastly, and perhaps most convincingly, we showed that the cultural proxy – i.e., female LFP in the country of ancestry in 1950 – had no explanatory power for the wages received by the second-generation women. Thus, if there were human capital differences that varied systematically by country of ancestry, these were not showing up in wages which robs the unobserved human capital story of most of its plausibility. Our analysis concluded that it was unlikely that unobserved human capital was driving our results.

I conclude this discussion by reminding the readers that this literature is in its infancy. The key questions of how culture is determined, how it changes, how it influences and is influenced by institutions are only just beginning to be studied. A central, simple, yet powerful, insight from economics is that people respond to incentives. These incentives, however, are, at least in part, culturally (socially) determined and thus tend to be self-perpetuating. Nonetheless, the consequences of innovations are often unpredictable and new information and learning can challenge and test societal beliefs, forcing them to evolve.[8] Indeed, in a constantly changing world such as ours over the last two hundred years, the view of culture as static is particularly mistaken.

References

Bisin, A. and T. Verdier (2000), "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits", *Quarterly Journal of Economics* 115, 955–88.

Chong, A, S. Duryea and E. La Ferrara (2007), "Soap Operas and Fertility: Evidence from Brazil", working paper.

Fernández, R. (2007), "Women, Work, and Culture", *Journal of the European Economic Association*, 5(2–3), 305–32.

Fernández, R. (2008) ,"Culture and Economics" in S. N. Durlauf and L. E. Blume, New Palgrave Dictionary of Economics, 2nd edition, Palgrave Macmillan, Basingstoke and New York.

Fernández, R. (2008), "Culture as Learning: The Evolution of Female Labor force Participation Over a Century", working paper.

Fernández, R. and A. Fogli (2006), "Fertility: The Role of Culture and Family Experience", *Journal of the European Economic Association* 4, 552–61.

Fernández, R. and A. Fogli (2009), "Culture: An Empirical Investigation of Beliefs, Work, and Fertility", *American Economic Journal: Macroeconomics*, 1 (1).

Guiso, L., P. Sapienza and L. Zingales (2006), "Does Culture Affect Economic Outcomes?", *Journal of Economic Perspectives* 20(2), 23–48.

Hanushek, E. and D. Kimko (2000), "Schooling, Labor-Force Quality, and the Growth of Nations", *American Economic Review* 90(5), 1184–1208.

---

[8] Cultural change can come as a result of learning over time (as in Fernández [2008]) who uses a model of intergenerational learning to explain why the change in female LFP over 100 years looks S-shaped), or because of the popular diffusion of alternative paradigms (as in La Ferrara et al.'s [2007] evidence that the portrayal of small families in soap operas may have influenced fertility in Brazil) or because of optimizing decisions on the part of parents (as in Bisin and Verdier's [2000] model of cultural transmission.

The Intergenerational Transmission of Attitudes

Thomas Dohmen*, Armin Falk**,

David Huffman*** and

Uwe Sunde****

Introduction

Understanding persistent differences in economic performance and material well-being across countries is (and has always been) a central concern of economists and policy makers. Modern economic theories usually model economic outcomes, on the individual level as well as on the aggregate level, as the result of rational decisions made by individuals. Individuals make optimal choices given their attitudes – preferences and beliefs about the environment and the behavior of others – and also their constraints, in order to maximize their well-being or whatever their objective in life is. The potential for differences in attitudes and "culture" to explain cross-country differences in economic outcomes is a long-standing idea in political economy and sociology, as evidenced by the work of Max Weber in the early 20th Century. For conceptual and technical reasons, however, economists have traditionally concentrated on investigating the role differences in constraints play for economic outcomes.

One reason why economists have only recently begun to study the link between attitudes and economic phenomena is that traditionally they have been skeptical about the reliability of measuring attitudes on a large scale using subjective measures, such as survey responses. The typical problems associated with subjective survey responses include a high level of

abstraction, dependence of responses on context, as well as biased responses due to considerations about the social desirability of the answers to certain questions. Without (economic) incentives to answer attitude questions accurately and truthfully, and with only a qualitative response scale on which to indicate attitudes, such measures could give very noisy or even misleading results, raising doubts about how meaningful such subjective measures are. Another, conceptual reason for the reluctance in investigating the role of attitudes is that economists regard preferences, and often also beliefs, as exogenously given by nature and immutable in order to be able to make sensible, non-trivial predictions about individual economic behavior. As a consequence, fairly little is known about the determinants of attitudes and their distribution within and across countries.

An increasing body of evidence points to stark differences in the attitudes of individuals, both within and across countries, however, with important implications for economic performance. For example, recent research shows that attitudes, such as the level of trust in a country, play an important role in explaining differences in income growth (see Knack and Keefer 1997). High levels of trust, measured by responses to survey questions asking individuals how much they trust other people, are associated with faster growth. This appears to reflect a causal impact of trust on economic activity rather than the other way around (see, e.g., Algan and Cahuc 2007). There is less research relating other attitudes to macroeconomic outcomes, due mainly to a lack of data, but risk attitudes are also likely to be very important. A recent study comparing the US and Germany shows a large difference in risk attitudes, with Germans being less willing to take risks (Naef et al. 2007). Risk attitudes are known to determine important behaviors at the individual level, such as investment in risky assets, and risky occupational choice, such as entrepreneurship (e.g., Dohmen et al. 2005; Bonin et al. 2007), and migration (Jaeger et al. 2009). Thus, greater willingness to take risks in a population is likely to pay off in the long run, with higher rates of investment and innovation, and a more efficient allocation of resources.

* Professor of education and the labour market and director of the Research Centre for Education and the Labour market (ROA) at Maastricht University.
** Professor for economics at the University of Bonn and director of the Bonn Laboratory for Economic Experiments..
*** Assistant professor for economics at Swarthmore College.
**** Professor for economics at the University of St.Gallen.

In light of this evidence and the undisputable role of attitudes in economic decision making, economists are just beginning to depart from the practice of treating attitudes as a black box. Recently, new approaches and instruments have been developed that allow researchers to reliably elicit attitudes in different domains of economic decision making. This, in turn, provides new ways for studying the process through which attitudes are formed, and investigating the determinants of attitudes provides a promising way of bringing culture back on the agenda of mainstream economics.

The measurement of attitudes: risk and trust

As interesting as it might be to find that cross-country differences in economic well-being are related to differences in trust or risk attitudes based on subjective survey responses, this finding would not be sufficient to convince a skeptical economist of the role these attitudes play. The reason is that survey responses could proxy for a host of other variables related to economic growth rather than actually measuring trust or risk attitudes. Hence, the construction of reliable, objective measures of attitudes in the population constitutes a necessary condition for qualifying and quantifying the role of attitudes and attitude transmission across generations for economic outcomes.

Recent research uses techniques from experimental economics to try to assess the validity of attitude measures. The idea behind this approach is to have subjects answer survey questionnaires that include standard attitude questions, and then have them participate in carefully constructed choice experiments that involve real economic payoffs. The experiments aim to isolate a specific attitude, for example towards trust or risk, and rule out other motives by design. For example, an experiment done with a representative sample of about 450 Germans found that a set of three standard trust questions reliably predicted trusting behavior in the so-called trust game (Fehr et al. 2003). One of our own studies involved an experiment about risk taking with a different representative sample of almost 500 Germans. In this study, it turned out that a simple question asking people to rate their general willingness to take risks on an eleven-point scale reliably predicted their willingness to play risky lotteries involving real money (Dohmen et al. 2005). This evidence lends credence to the interpretation of previous studies, in which the relationship between attitude measures and growth is actually based on the impact of specific attitudes.

Culture: intergenerational transmission of attitudes

If attitudes play such an important role for economic decision making and individual and aggregate outcomes, how are attitudes determined? How do attitude differences across regions or countries come to exist, and how do they persist over time? One explanation is that, due to some differences in the environment, or because of major historical events, attitudes diverge between nations at some point in time. These differences in attitudes are then passed on from one generation to the next, through a channel of parents inculcating children with their own values and attitudes (Tabellini 2008), through imitation of parents by children, or even, perhaps, through genetics. Departing from the usual practice in economics of taking preferences, or attitudes, as given, a wave of recent theoretical models have made attitudes endogenous by explicitly assuming that parents have a preference for endowing children with attitudes similar to their own and exert effort in order to shape their children's endowments (see, e.g., Fernandez 2007). This preference also affects parents' marriage choices, causing them to seek out a spouse with similar attitudes, in order to avoid distortions in the child's attitudes. Parents pass on their attitudes to their children, perpetuating differences in economic behavior between populations with different attitudes or cultures. Some models also allow for an influence of other role models in the surrounding environment, in addition to, or instead of, the parents, which tends to reinforce and perpetuate regional differences in attitudes (see Bisin and Verdier 2000). Moreover, there might be an interaction between parents' effort in inculcating their children with certain attitudes, and the economic or technological environment (see Doepke and Zilibotti 2008).

While assuming a transmission of attitudes from parents to children is plausible and allows theoretical models to explain an important set of facts, there has been little systematic empirical evidence on the intergenerational transmission at the level of attitudes. Some of our recent work tests whether children end up with similar trust, and similar risk attitudes as their parents (Dohmen et al. 2006). We use the experimentally validated survey measures discussed above, for a sample of more than 3,000 children and their parents. The data are drawn from the 2003 and 2004

Figure

CHILD'S ATTITUDES TOWARDS RISK AND TRUST
AS FUNCTIONS OF PARENTAL ATTITUDES

Child's risk attitude (mean) and weighted regression line

Child's trust attitude (mean) and weighted regression line
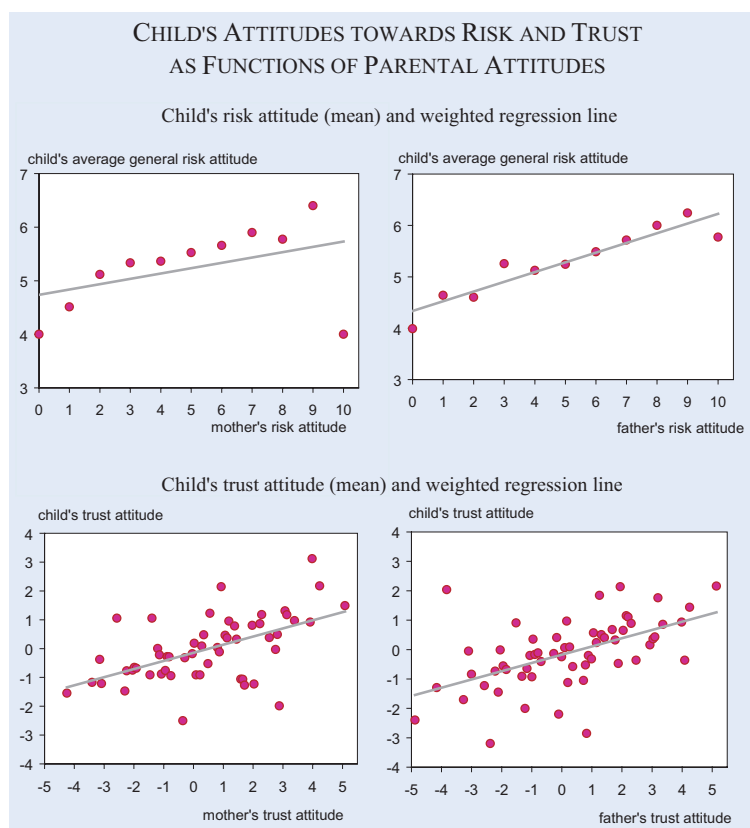
Notes: The upper graphs in the figure show children's average self-reported willingness to take risks, from 0 (completely unwilling) to 10 (completely willing), for a given willingness to take risks on the part of the parent. The bottom graphs in the figure show children's average principal component "trust", a summary measure of the level of agreement with three different statements about trustworthiness of people in general, for a given principal component "trust" of the parent.

ilar plot for trust. The child's mean trust attitude, measured by a standardized principal component obtained from the responses to three trust-related survey questions, is plotted against the mother's (left graph) and father's (right graph) trust attitudes. Again, there is a significant and sizable correlation. However, in the context of trust, the mothers appear to play a somewhat more important role for the children's attitudes than fathers do.

These correlations are essentially unchanged when controlling for similarity across generations in personal or environmental characteristics, showing that the relationship between the attitudes of parents and children does not work through these indirect channels. The relationship between parents and child is also present using various additional questions about the same attitudes, with different scales and framings. Additional robustness checks show that the intergenerational cor-

waves of the SOEP, a representative panel survey of the adult population living in Germany. Our approach involves using direct measures of attitudes for individuals as well as for their parents. We control for detailed characteristics of parents, such as income and education, which are important for addressing potential issues of omitted variables (see Fernandez 2007).

The main findings are illustrated in the Figure. The upper two panels of the Figure depict the child's average general risk attitude, measured on a scale from 0 to 10, conditional on the risk attitude of the mother (left graph) and father (right graph). The scatter plot, as well as the regression line, which is weighted by cell-size, provide a strong indication that parents who state they are more willing to take risks have children with similar attitudes. The correlation between children's risk attitudes with those of both mothers' and fathers' is statistically significant, quantitatively important when looking at outcomes like income or regional mobility, and of about similar size for both parents. The lower panel of the figure provides a sim-

relation is not explained by parents and children collaborating on survey answers, by similar scale use, by reverse causality from children to parents, or by parents and children living in the same geographic region. The results also indicate that children are not just similar to their parents in terms of overall attitudes, or a diffuse disposition towards risk-taking and trusting, but that children are similar to their parents in an even more precise sense. For example, when investigating the transmission of risk attitudes, controlling for the child's and the parents' trust, children's risk attitudes are strongly and significantly associated with those of their parents, but not, or only very weakly, with their trust. The results for more detailed, context-specific measures of risk-taking are similar, with parental attitudes in a particular context (e.g., health, financial matters, career and leisure) being the strongest predictor of the child's attitude in that context. The same holds for trust, where parents' attitudes in a given context are the best predictor of a child's attitudes in that same context.

In addition to this evidence for the direct transmission of attitudes from parents to children, the intergenerational transmission mechanism of attitudes might be reinforced through two additional channels. One is positive assortative mating of parents. If parents marry partly on the basis of attitudes to be transmitted to the child, a child who has one parent with a given attitude is likely to have a second one with that attitude as well. Assuming that both mothers and fathers matter for a child's attitudes, positive assortative mating is an implication of models that assume parents have a preference for children with attitudes similar to their own; in this case parents have an incentive to find similar partners, in order to avoid distortions in the transmission of attitudes to their children (see, e.g., Ichino and Maggi 2000; Bisin and Verdier 2001; and Bisin et al. 2004). Our findings indeed suggest that parents tend to marry individuals with similar trust and risk attitudes.

Another mechanism involves other role models in the environment influencing a child's attitudes, in addition to the parents. This mechanism arises because a child's attitudes are assumed to be susceptible to socialization, and would tend to reinforce regional or ethnic differences. We therefore also investigated whether a child's attitudes are related to the prevailing attitudes in the local geographic region. The findings indicate a role for environment, in that a child's attitudes are similar to the prevailing attitudes in the local geographic region, even controlling for parental attitudes. Of course, these two alternative reinforcing mechanisms interact provided that parents sort themselves into neighborhoods with a population that has attitudes similar to their own.

Nurture or nature?

An important question concerns the precise channel through which attitudes are transmitted. It is still an open question whether attitude transmission works through nurture (i.e., deliberate inculcation or imitation) or nature (i.e., physiological channels like genetics), or both. There is evidence from twin studies that risk attitudes are genetically inheritable. Recent findings suggest that about 20 percent of the variation in individual risk attitudes might be explained by genetic differences (see, e.g., Cesarini et al., 2009). By now, several pieces of evidence suggest that nurture must play also at least some role. For example, Dohmen et al (2006) find that family structure has an impact on children's attitudes. Single-children are

more similar to their parents in terms of attitudes than children with siblings are, which is hard to explain with a purely genetic mechanism. It is consistent with single-children receiving more undivided attention and socialization from the parents, however. Likewise, it appears difficult to reconcile the influence of the attitude distribution in the region of residence with a purely genetic transmission mechanism.

Policy implications

To conclude, it appears that the data are strongly consistent with a transmission of economically relevant attitudes from parents to children. An important area for future research is investigating the relative strength of the different transmission channels as well as their interactions. How different parental characteristics and family structures matter for the transmission of attitudes from parents to children has important implications for the potential of policy interventions.

Evidence for the transmission of attitudes from parents to children is highly relevant for understanding why there is a strong persistence in economic outcomes across generations for different families, dynasties and even countries. There is a large literature studying social mobility within countries, which documents substantial correlations between parents and children in terms of income, wealth, education and occupation. Transmission of attitudes could be one mechanism underlying such correlations: one reason that children may end up with outcomes similar to their parents may be that they inherit similar attitudes and thus make similar economic choices. Trust and risk attitudes are both relevant for the types of outcomes that are typically correlated between parents and children, such as wealth accumulation and occupational choice. Other attitudes are also likely to be important, however, for example patience.

A role for attitudes in explaining intergenerational correlation in economic outcomes points to the importance of the child's home environment and parental attitudes as policy levers for addressing lack of social mobility. Interfering with the ways that a family raises its children is of course controversial and problematic from a policy perspective.

Nevertheless, recent evidence points to important correlations between parental characteristics such as education and their attitudes, with more educated

parents being more trusting and more willing to take risks (see, e.g., Dohmen et al. 2006). Recent evidence also documents that risk attitudes and patience are related to cognitive skills (see Dohmen et al. 2007). Evidence by Carniero et al. (2007) shows that influencing mother's education might be a useful way to affect child outcomes like test scores, grade repetition and obesity. Hence, this appears highly relevant as a potential way to affect the transmission of attitudes among certain population groups, regions or countries and to thus improve their economic perspectives. The results cited above suggest one mechanism through which policies focused on parental education could affect a child's outcomes, namely through their cultural transmission: the transmission of economically relevant attitudes from parents to children.

References

Algan, Y. and P. Cahuc (2007), "Social Attitudes and Economic Development: An Epidemiological Approach", IZA *Discussion Paper* No. 2935.

Bisin, A. and T. Verdier (2000), "'Beyond the Melting Pot': Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits", *Quarterly Journal of Economics,* 115(3), 955–88.

Bisin, A. and T. Verdier (2001), "The Economics of Cultural Transmission and the Dynamics of Preferences", *Journal of Economic Theory,* 97(2), 298–19.

Bisin, A., G. Topa and T. Verdier (2004): "Religious Intermarriage and Socialization in the United States," *Journal of Political Economy,* 112(3), 615–64.

Bonin, H., T. Dohmen, A. Falk, D. Huffman and U. Sunde (2007) "Cross-Sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes" *Labour Economics,* 14(6), 926–37.

Carneiro, P., C. Meghir and M. Parey (2007), "Maternal Education, Home Environments and the Development of Children and Adolescents", *CEPR Discussion Paper* DP6505. Link: www.cepr.org/pubs/dps/DP6505.asp

Cesarini, D., C.T. Dawes, M. Johannesson, P. Lichtenstein and B. Wallace (2009), "Genetic Variation in Preferences for Giving and Risk-Taking", *Quarterly Journal of Economics,* in press.

Doepke, M. and F. Zilibotti (2008) "Occupational Choice and the Spirit of Capitalism", *Quarterly Journal of Economics,* 123(2), 747–93.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp and G.G. Wagner (2005), "Individual Risk Attitudes: New Evidence From a Large, Representative, Experimentally-Validated Survey", *IZA Discussion Paper* No. 1730.

Dohmen, T., A. Falk, D. Huffman and U. Sunde (2006), "The Intergenerational Transmission of Risk and Trust Attitudes", *IZA Discussion Paper* No. 2380.

Dohmen, T., A. Falk, D. Huffman and U. Sunde (2007), "Are Risk Aversion and Impatience Related to Cognitive Ability?", *IZA Discussion Paper* No. 2735.

Fehr, E., U. Fischbacher, B.v. Rosenbladt, J. Schupp and G.G. Wagner (2003), "A Nation-Wide Laboratory – Examining Trust and Trustworthiness by Integrating Experiments in Representative Surveys", *IEW Working Paper* No. 141.

Fernandez, R. (2007): "Culture and Economics", in S. Durlauf and L. Blume (eds.), *New Palgrave Dictionary of Economics, Second Edition,* Palgrave Macmillan (Basingstoke and New York).

Ichino, A. and G. Maggi (2000), "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm," *Quarterly Journal of Economics,* 115(3), 1057–090.

Jaeger, D., T. Dohmen, A. Falk, D. Huffman, U. Sunde and H. Bonin (2009), "Direct Evidence on Risk Attitudes and Migration", *Review of Economics and Statistics,* in press.

Knack, S., and P. Keefer (1997), "Does Social Capital Have an Economic Payoff, A Cross-Country Comparison", *Quarterly Journal of Economics,* 112(4), 1251–88.

Naef, M. E. Fehr, U. Fischbacher, J. Schupp and G.G. Wagner (2007), "Decomposing Trust: Explaining National and Ethnical Trust Differences", *IEW Mimeo.*

Tabellini, G. (2008), "The Scope of Cooperation: Values and Incentives", *Quarterly Journal of Economics,* 123(3), 905–50.

The Effect of Culture on the Functioning of Institutions: Evidence from European Regions

Mariaflavia Harari* and
Guido Tabellini**

Culture: the missing link?

An influential line of research starting with North (1981) and continuing with Acemoglu, Johnson and Robinson (2001) has stressed the importance of distant history as one of the main determinants of current economic development. In particular, economic backwardness has been identified as a by-product of a past history of despotism or slavery. But what is the channel through which distant history shapes current economic outcomes?

A widespread interpretation is "institutions". However, economic and policy outcomes are often much more persistent than the institutions which are supposed to shape them. Institutions, in fact, may change abruptly, and yet leave economic outcomes nearly unaffected. Consider for example transitions from autocracy to democracy and vice versa: while coups seem to bring about a pronounced deterioration in economic outcomes, democratic transitions are not associated with comparably large improvements. On the other hand, consider a country where political institutions have been the same for centuries. Within that country, economic and policy outcomes may diverge across regions; these differences can be traced back to different regional histories in the past, whose legacy persists in spite of identical political and legal institutions. In Italy, for instance, it is well known that hospitals, courts, schools and local governments are much less efficient in the South than in the North.

* Research assistant at Bocconi University.
** Rector and professor of economics at Bocconi University.

Clearly, the persistence in institutional outcomes cannot be fully explained by looking at formal institutions only: something else, besides institutional inertia, must account for this legacy of history.

Political economy explains this persistence with reference to the redistributive conflicts and economic incentives of political and economic élites who shape institutions so as to preserve their *status quo* rents. However, economic incentives alone fail to explain several institutional failures: why does a civil lawsuit in Southern Italy last three times as much as in Northern Italy, in spite of identical legislation and incentives, and similar resources? Why do rational voters keep voting for corrupt politicians? More generally, why is it possible to overcome the collective action problem, the core of almost any form of political participation, in some situations, but not in others? The problem is that while individual incentives play a crucial role in most economic situations, this is not the case in several political situations, in which aggregate outcomes reflect the actions of many atomistic individuals. Moreover, inside government organizations, incentives are generally weak, due to lack of competition.

Besides purely economic incentives, we should devote our attention also to other relevant drivers of individual behavior, such as morality. One's belief about the behavior of other people, for example, exerts a strong influence on voters' demands and on citizens' participation; one's perception of what is right or wrong determines the extent of moral hazard inside public organizations. In other words, "culture", by which we mean primarily normative values on what is "right" or "wrong" and how one "ought" to behave in given circumstances, is likely to affect the functioning of institutions.

Values are largely transmitted vertically, from one generation to the next, in a conservative mechanism which takes place mostly within the family, rather than across unrelated individuals. As suggested also by Roland (2004) and Guiso et al. (2006), such slow moving values can indeed be the "missing link" through which distant political history influences the functioning of current institutions and current devel-

opment, and can explain the persistence of institutional outcomes.

Two stylized facts, discussed in detail in Tabellini (2008), suggest that this is indeed the case. First, it makes sense to talk about the quality of government as a general feature of countries: some countries are governed "well" in many policy areas and others are governed "poorly" also in many areas. This is confirmed by cross country data (Tabellini 2008), showing that government failures and successes are correlated across many different policy dimensions. That policy distortions and government inefficiencies are often clustered together seems to suggest that they may have a common cause. Second, culture is slow moving: exploiting data on third generation immigrants to the US, Tabellini (2008) shows that current values, extrapolated by the World Value Surveys opinion polls, reflect features of the country of origin of the respondents' ancestors. Similar results have also been obtained by Guiso et al. (2006). This article reviews the main results in these recent contributions, drawing on Tabellini (2008; 2009).

Values and institutional outcomes

The first question to be addressed is which cultural traits are especially important in explaining institutional outcomes, and in particular which ones are likely to induce well-functioning institutions.

One key contraposition is between values consistent with "limited" as opposed to "generalized" morality; the distinction concerns the scope of application of norms of good conduct: within a narrow group with which the individual identifies, or towards everybody. In hierarchical societies, codes of good conduct tend to be confined to small circles of related people, outside which opportunistic and selfish behavior is regarded as natural and morally acceptable. In his case study of life in a rural village in Southern Italy (*The Moral Basis of a Backward Society* 1958), political scientist Edward Banfield calls this attitude "amoral familism", referring to the fact that principles of good and evil are applied inside the family only. In modern democratic societies, on the other hand, abstract rules of good conduct apply to many social situations. As argued by Max Weber ([1905] 1970), the emancipation of the individual from feudal arrangements has typically been associated with a diffusion of generalized morality. Yet, the distinction generalized/limited morality is still relevant to-

day, in order to understand cultural differences between different countries or regions.

The idea that generalized morality leads to better collective outcomes has a long history in political science and has recently been gaining grounds in economics as well. The relevance of culture and morality to the functioning of institutions has been stressed amongst others by Landes (1998) and Platteau (2000). A related idea is that a civic culture and a well-educated population are an important prerequisite to a well functioning and stable democracy, as highlighted, inter alia, by Glaeser et al. (2006), and Persson and Tabellini (2006).

Generalized morality norms are likely to instill confidence and respect for abstract principles such as the rule of law, the respect for individual rights and for democratic procedures. Hence they are likely to induce reciprocal cooperation and discourage free riding, with virtuous consequences not only for the economic behavior of individuals but also for their participation in the political and administrative life of their local communities. The latter is crucial to ensure an adequate provision of local public goods and to monitor political representatives or local administrators. Thus, generalized morality is conducive to well functioning institutions through at least three channels: law enforcement is easier because citizens are more likely to be law abiding (eg., less likely to cheat on taxes); bureaucrats are less corrupt; voters expect and demand higher standards of behavior from politicians, monitor public administrators more effectively and are more inclined to vote based on general social welfare rather than personal benefit criteria.

On the other hand, in societies where limited morality is prevalent, individualism is mistrusted; the role for the state, as well as for parental education, is a repressive one: good behavior is deemed to result from coercion, not from internalization of the values of society. Such coercive cultural environments stifle individual initiative and entrepreneurship as well as cooperation within a group, with adverse consequences on economic development. Aghion et al. (2009) suggest that low levels of social capital induce a strong demand for government regulation; conversely, a high level of regulation discourages civic behavior, since, when entrepreneurship is restricted through regulation, investment in social capital may not pay off. In their model, values and institutions mutually affect each other: when people expect to live in an "uncivic" community, they expect high levels of regulation and

corruption, and do not invest in social capital; their expectations are fulfilled and the society ends up trapped in a "bad equilibrium" with high levels of corruption and regulation.

The empirics

At the empirical level, it is possible to construct measures for the diffusion of norms of generalized vs limited morality at an aggregate level exploiting the attitudes revealed by The World Value Surveys opinion polls (Inglehart et al. 2000). A variable extensively used in the economic literature is *trust*, defined as the fraction of respondents believing that "most people can be trusted". This has been interpreted in two alternative, but not mutually exclusive ways: as belief about the behavior of others, and as an indicator of moral values and trustworthiness. A moral interpretation has been advocated amongst others by Ulsaner (2005), who has shown that *trust* is a very persistent individual feature correlated with charitable contributions and volunteering, and by Glaeser et al. (2000).

In a similar way we can construct the variable *respect*, as the fraction of respondents who claim to consider "tolerance and respect for others" as an important quality which children should be encouraged to learn at home. These two cultural traits encapsulate what earlier studies have called "social capital". One of the possible interpretations of these variables is as measures for individual values related to generalized morality; conversely, lack of trust and lack of respect for others' actions are typical of hierarchical, coercive societies dominated by limited morality. The above-mentioned variables can be considered in isolation or combined to construct a summary measure of the cultural traits favorable to the good functioning of institutions.

In cross-country data, *trust* has been shown to be correlated both with favorable economic outcomes and with indicators of well functioning institutions. The latter are measured by perceptions of bureaucratic quality and indicators of property rights protection. Although this correlation should not carelessly be interpreted as evidence of a causal link, the explanatory power of culture in cross-country regressions is remarkable. Consider, for example, the Netherlands and Italy, two countries at comparable levels of development but respectively close to the top and bottom of the distribution of the governance indicator amongst the OECD countries: according to the spec-

ification of Tabellini (2008), over half their difference in governance may be explained by differences in the variables *trust* and *respect*. Aghion et al. (2009) have also highlighted a strong negative correlation at the cross-country level between *trust* and government regulation.

The evidence from European regions

Drawing inferences from cross country data is problematic: countries diverge in a large number of economic and political dimensions, so that the differences in economic outcomes may be due to factors other than culture and not explicitly included in the specification (omitted variables). Moreover, institutions and culture are likely to interact with each other to determine economic development, and working with cross-country data it is hard to disentangle the two effects. Both shortcomings can be overcome by exploiting variation in economic outcomes at the subnational level, focusing on the role of culture as a channel of historical influence within, rather than across, countries.

The importance of social capital within countries had been stressed for the first time by political scientist Robert Putnam in his seminal work on Italy's civic divide (1993), in which the diverging performances of Italian regions are attributed to their different civic traditions. In particular, Putnam points to the free city-state experience during the Middle Ages and the culture of independence which it fostered. Guiso et al. (2008) test these conjectures investigating the historical origins of social capital across Italian cities and estimate that at least 50 percent of the North-South gap in social capital is due to a lack of a free city-state experience in the South. De Blasio and Nuzzo (2006) also conduct a study on differences in social capital across Italian regions.

Values can influence regional economic development through a variety of channels: from the functioning of the public administration, to criminal activities in the region, to behavior inside private organizations. For example, Ichino and Maggi (2000) investigate moral hazard inside a large Italian bank with branches in different regions, finding that absenteeism and misconduct episodes are substantially more relevant in the South.

Tabellini (2009) considers a sample of 69 regions located in a small number of homogeneous European
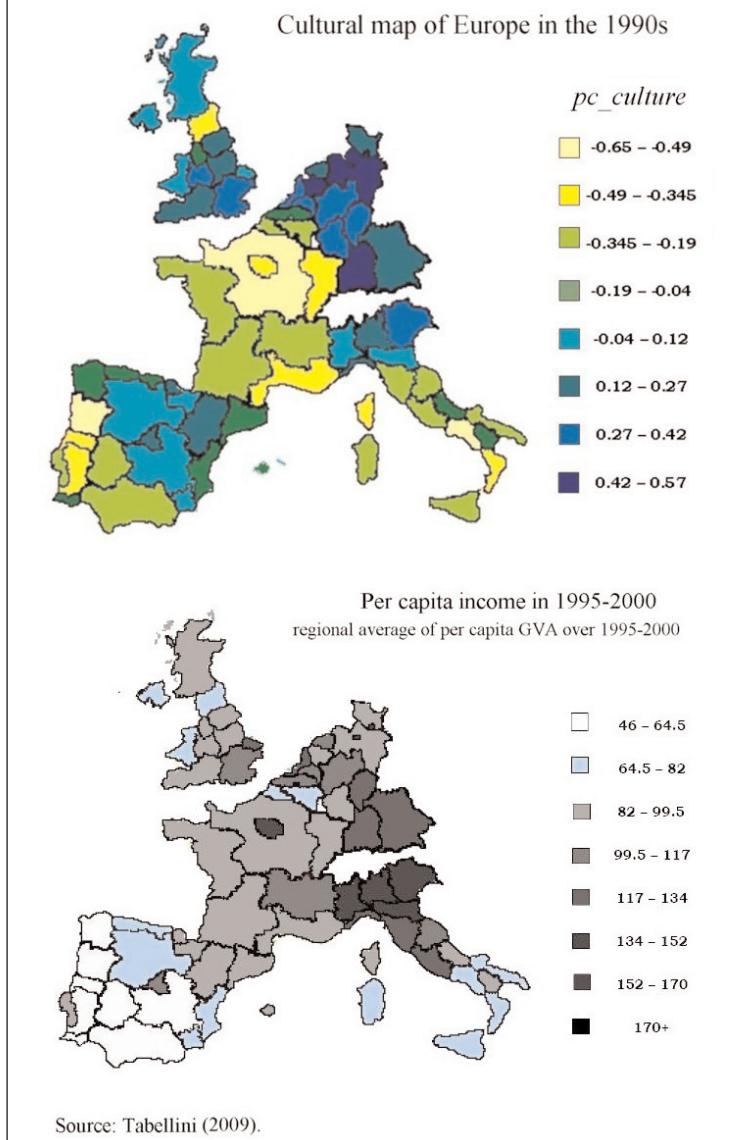
Figure 1

Cultural map of Europe in the 1990s

pc_culture

- -0.65 – -0.49
- -0.49 – -0.345
- -0.345 – -0.19
- -0.19 – -0.04
- -0.04 – 0.12
- 0.12 – 0.27
- 0.27 – 0.42
- 0.42 – 0.57

Per capita income in 1995-2000
regional average of per capita GVA over 1995-2000

- 46 – 64.5
- 64.5 – 82
- 82 – 99.5
- 99.5 – 117
- 117 – 134
- 134 – 152
- 152 – 170
- 170+

Source: Tabellini (2009).

Figure 1 shows the regional averages of the summary variable *pc_culture*[3] and regional per capita output (measured as average gross value added over the period 1995–2000). The regional pattern of culture is strikingly similar to that of per capita output. In particular, Germany, England and Northern Italy tend to have high per capita output and positive cultural indicators, while Southern Italy, Portugal and Southern Spain fare worse on both counts. The correlation, however, is not perfect: in particular, France has high per capita income in spite of cultural traits which are a priori less favorable to economic development.

This correlation could merely reflect the influence of other common determinants, such as national institutions, education or past levels of economic development. But a deeper analysis, details of which can be found in Tabellini (2008; 2009), reveals that this is not the case: once we control for country-fixed effects and other regional variables, including school enrolment in 1960, urbanization in 1850 and literacy in 1880, culture and per capita output still appear positively correlated, as we can see from the interpolant in Figure 2.

According to this specification, a difference in culture of 50, such as between Lombardy and a typical region in Southern Italy, is predicted to be associated with a difference in GDP per capita of about one third of the EU average; this would be tantamount to accounting for almost half of the observed income difference between Lombardy and Southern Italy.

countries.[1] The countries considered share the following peculiarity: although formal and legal institutions at the national level have been the same for 150 years or more, there is a variety of regional political histories within each country. The key idea is to compare current economic outcomes in European regions, controlling for country-fixed effects, in such a way that the effect of the common national institutions is removed. At this point, it will be possible to determine to what extent differences in economic performance can be explained by differences in culture, which in turn are induced by different regional histories.[2]

---

[1] Belgium, France, Italy, Netherlands, Portugal, Spain, UK and West Germany.
[2] Beugelsdijk, and von Schaik (2001) also study the correlation between social capital and per capita output in European regions, but do not attempt to link social capital to history nor to account for endogeneity.

[3] *Pc_culture* is constructed to include four cultural variables extrapolated from the World Value Surveys: *trust, respect, control* and *obedience*. *Control* captures one's belief in individual effort and initiative, which is correlated with generalized morality, while *obedience* captures a cultural emphasis on hierarchy as opposed to egalitarianism, and is correlated with limited morality. *Pc_culture* is strongly positively correlated with *trust* and *respect* and represents an overall indicator of the cultural traits favourable to well-functioning institutions.

Figure 2

Correlation between output and culture,
including country fixed effects and regional controls.
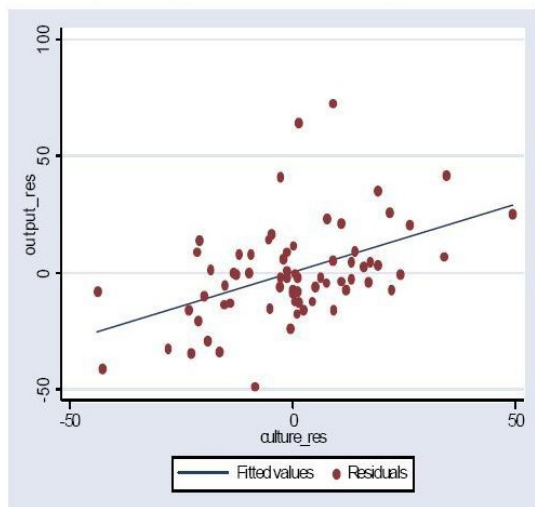
Source: Tabellini (2009).

This strong correlation between culture and economic development, however, does not by itself imply a causal link which goes from culture to development. The key difficulty in estimating a causal effect of culture is that it is endogenous to economic development and poses problems of reverse causality: not only does "good culture" induce "good institutions" and development, as we wish to show, but "good institutions" induce "good culture" as well, as underlined by the so-called modernization theory (e.g., Inglehart and Baker 2000). In other words, the correlation we have highlighted may be due in whole or in part to the fact that economic development induces values consistent with generalized morality, rather than the other way around. In order to rule out this reverse causality, we have to rely on some exogenous source of variation in culture, resorting to the technique of instrumental variable regression ("two-stage regression").

Instead of culture, we consider an "instrument" for culture: the quality of political institutions ruling the region in the distant past (i.e., before the countries considered became unified). The crucial identifying assumption is that past institutional quality affects current values but has no direct impact on current development. Under this hypothesis, if we find any impact of past institutional quality on current development, this effect must necessarily occur through culture, according to the causal mechanism: early political institutions => regional cultural traits => current development. The reverse causality problem is then ruled out, since current development can by no means have affected past political institutions.

These identifying assumptions require a careful evaluation. A first-stage regression of current culture on past institutional quality, measured by considering checks and balances on the regional executives between 1600 and 1850, reveals a robust correlation: past institutional quality indeed seems to shape current regional cultural traits. This is consistent with the idea that regions where despotic governments exploited citizens are likely to have inherited a culture of mistrust and limited morality, whereas regions ruled by republican regimes, where participation is promoted and the rule of law is respected are likely to have inherited a culture of generalized morality. It is also interesting to note that, on the other hand, no significant correlation is found between current culture and urbanization in 1850. This supports the identifying assumption that contemporaneous cultural traits do not just reflect economic development in previous centuries, but rather the political environment in which previous generations used to live.

The second assumption, i.e., that past institutional quality has no direct correlation with current development, appears reasonable, although it cannot be tested directly. What we are assuming is that, once we control for past economic development as measured by literacy in 1880 and urbanization in 1850, institutions in place before the unification have no impact on current development. This is tantamount to assuming that 150 years of unification and national policies have made up for those regional differences in development which are due to the inheritance of previous regimes. Under these assumptions, the remaining differential in current development across regions is attributed to differences in culture.

Resorting to history as an instrument for culture, the estimated impact of culture on output is still positive and significant, and appears to be larger compared to the simpler specification. In other words, the cross-regional variation in culture that can be attributed to history appears to be more strongly correlated with development compared to the overall measures of culture. This result, however, should be considered with caution: the two-stage regression could yield higher estimates due to the invalidity of the instruments or measurement error.

These findings are very robust to adding other regressors, alternative measures of values, or alternative estimation strategies, as discussed extensively in Tabellini (2008; 2009).

Figure 3

EFFECT OF RAP ON PREFERENCE VOTES BY TYPE OF ELECTORAL DISTRICT

- Effect of RAP in worst district
- Effect of RAP in best district
- Average effect of RAP

% votes lost

Constraints on Exec 1600–1800 | Trust | Respect

Source: Nannicini et al. (2009).

so by Parliament. Prosecutors' requests to proceed with criminal investigations of elected representatives (called RAP from here on) are public knowledge. We draw on data on Italian regional elections before 1993, when voters could express preference votes over individual candidates to investigate how voters react to RAP on incumbents in different electoral districts.[4] Controlling for district and legislature fixed effects, as well as individual features of the incumbent, we find that voters in regions with a better political history, or with higher values of trust and respect, punish incumbents more, in terms of preference votes, for having received a RAP. In the regions with the lowest values of *trust* and *respect* votes received by an incumbent receiving a RAP increase slightly, although generally by a negligible amount; conversely, in districts with high values of *trust* and *respect* votes received drop by as much as 20 percent to 35 percent, depending on the specification, as summarized in Figure 3.

The evidence presented so far supports the idea that culture is a long-run determinant of output, measured in levels. But does culture affect also regional economic growth in the short run? The evidence suggests that it does. The analysis sketched so far can be replicated, using as a dependent variable regional output growth over the period 1977–2000. Again, we find that the component of regional values explained by distant political history has a positive and significant association with regional growth. In particular, according to the estimated coefficient, if Southern Italy had the same culture as Lombardy, its average yearly growth rate would have been higher by almost 0.5 percent (Tabellini 2009).

Summing up, all the instrumental variable estimates discussed portray a remarkably consistent and robust picture: first, past political institutions and low literacy rates left a mark on regional culture; second, this cultural legacy of history is an important determinant of current economic performance. Moreover, the data cannot reject that past political institutions and literacy rates of previous generations influence economic performance only through culture.

A further interesting piece of evidence of how regional differences in culture can explain divergent institutional outcomes comes from the analysis of Italian voting behavior. In ongoing preliminary work, Nannicini et al. (2009) test the hypothesis that voters sharing norms of generalized morality demand higher standards of behavior on their elected representatives.

In Italy prosecutors cannot investigate elected representatives unless they first obtain authorization to do

Implications for research and policy

The evidence presented casts doubts on the primacy of formal institutions as determinants of economic development. Instead it suggests that culture is likely to interact with formal institutions influencing the incentives and the behavior of economic and political agents.

These findings suggest a promising research agenda. Attention should be devoted to how values are formed in the first place, why they reflect past institutions and how they are transmitted over time, following the line of research of Bisin and Verdier (2005) and Fernandez (2007). It would be equally interesting to investigate how values interact with formal institutions and incentives, in particular how do political outcomes (eg., targeted redistribution) shape cultural traits (eg., reinforcing group identity); and vice versa how culture influences political outcomes, through voters' behavior or in the formation of pressure groups.

---

[4] These data were also studied by Chang and Golden (2004), who however did not focus on heterogeneities across electoral districts.

If confirmed by future research, these findings also suggest a number of relevant policy implications for the regions of Europe and, in particular, Italy, which is characterized by a persistent North-South divide. The low labor productivity of economically backward regions may be linked to adverse cultural traits prevailing in those regions, implying that it is not likely to go away soon. Income transfers and public investment cannot be a solution, because they do not address the source of the problem. Instead, economically and culturally poor regions are likely to benefit from investments in education, from cheap sources of finance (to facilitate the emergence of local entrepreneurs), and from a decentralization of administrative and political powers, in order to stimulate the accumulation of social capital. These findings also reinforce the simple, but often neglected idea that regions with lower productivity ought to pay lower real wages: a single national wage concentrates unemployment in the poor regions (as it happened in Southern Italy and East Germany), self-perpetuating the adverse cultural features that might be at the root of the low labor productivity in these regions.

Moreover, this analysis suggests that the same policies may work differently in different regions, according to the different values prevailing in those areas. For example, active labor market policies are more likely to be abused in Southern Europe, where values do not to discourage free riding. Decentralization may lead to learning or to divergence: while some regions may benefit from decentralization, others may perform worse in a decentralized system. Attention should also be devoted to how different policy instruments influence values. More generally, an effort should be made in order to individuate policy instruments which fit the value system, taking into account how policies and values interact.

References

Acemoglu, D., S. Johnson and J. A. Robinson (2001), "The Colonial Origins of Comparative Development: An Empirical Investigation", *American Economic Review* 91, 1369–401.

Aghion, P., Y. Algan, P. Cahuc and A. Shleifer (2009), "Regulation and Distrust", mimeo, Harvard University.
Banfield, E. C. (1958), *The Moral Basis of a Backward Society,* The Free Press, New York.

Beugelsdijk, S. and T. von Schaik (2001), "Social Capital and Regional Economic Growth", *CentER Discussion Paper,* December.

Bisin, A. and T. Verdier (2005), "Cultural Transmission", in *The New Palgrave Dictionary of Economics;* Palgrave Macmillan, New York.

Chang, E., and M. Golden (2004) "Does Corruption Pay? The Survival of Politician Charged with Malfeasance in the Postwar Italian Chamber of Deputies", paper presented at the annual meeting of the American Political Science Association, Chicago, September.

De Blasio , G. and G. Nuzzo (2006), "The Legacy of History for Development: the Case of Putnam's Social Capital", mimeo, Bank of Italy.

Fernandez, R. (2007), "Culture as Learning: the Evolution of Labor Force Participation over a Century", *NBER Working Paper* no.13373.

Glaeser, E. L., D. I. Laibson, J. A. Scheinkman and C.L. Soutter (2000) "Measuring Trust ", *Quarterly Journal of Economics,* 115 (3), 811–46.

Glaeser, E. L., G. Ponzetto, and A. Shleifer (2006), "Why Does Democracy Need Education?", *NBER Working Paper* no.12128.

Guiso, L., P. Sapienza, and L. Zingales (2006), "Does Culture Affect Economic Outcomes?", *Journal of Economic Perspectives,* 20 (2), 23–49.

Guiso, L., P. Sapienza, and L. Zingales (2008), "Long Term Persistence", mimeo, University of Chicago.

Ichino, A., and G. Maggi (2000), "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm", *Quarterly Journal of Economics,* 15 (3), 1057–90.

Inglehart, R. et al. (2000), *World Values Surveys and European Values Surveys, 1981–1984, 1990-1993 and 1995–1997,* Ann Arbor, Michigan, Institute for Social Research, ICPSR version.

Inglehart, R. and W. E. Baker (2000), "Modernization, Cultural Change and the Persistence of Traditional Values", *American Sociological Review,* February, 65, 19–51.

Landes, D. (1998) *The Wealth and Poverty of Nations,* Abacus.

Nannicini, T., A. Stella, G. Tabellini and U. Troiano (2009), "Political Accountability and Social Capital", in progress, Bocconi University.

North, D. (1981), *Structure and Change in Economic History,* Norton, New York.

Persson, T. and G. Tabellini, (2006), "Democratic Capital: The Nexus of Political and Economic Change", *IGIER Discussion Paper,* Milan.

Platteau, J-P. (2000), "Institutions, Social Norms, and Economic Development", *Academic Publishers,* Amsterdam.

Putnam, R. D. (1993), *Making Democracy Work. Civic Traditions in Modern Italy,* Princeton.

Roland, G. (2004), "Understanding Institutional Change: Fast-Moving and Slow-Moving Institutions", *Studies in Comparative International Development,* 109–131.

Tabellini, G. (2008), "Institutions and Culture: Presidential Address", *Journal of the European Economic Association,* 6 (2–3), 255–94.

Tabellini, G. (2009), "Culture and Institutions: Economic Development in the Regions of Europe", *Journal of the European Economic Association,* in press.

Uslaner, E. M. (2005), "Trust as a Moral Value", in D. Castiglione, J.W. van Deth, and G. Wolleb, eds., *Handbook of Social Capital,* Oxford University Press.

Weber, M. ([1905]1970), *The Protestant Ethic,* George Allen and Unwin, London.

CULTURAL ECONOMICS

Bruno S. Frey*

Culture and Economics

The two terms "culture" and "economics" can be related in two quite different ways: The *first* approach is to study the relationship between two different sectors or spheres of society; the *second* approach is to study the arts with the help of economic analysis. The concept of "culture" as an *institution or as an organization supplying artistic services* used in the Economics of Art differs fundamentally from the concept that understands "culture" as an *attitude* or as a *way of behaving* used in the other papers in this collection. In order to illuminate the way the Economics of Art approaches a subject, the case of art museums is sketched.

Institutions of culture and the economy

The effect of *culture in an institutional sense on the economy* is extremely difficult to capture. What is, for instance, the effect on economic activity of having a theatre or a museum in a city? There is a great number of *"impact studies"* seeking to measure the effect of such institutions on firms located in a particular area. These studies, however, are based on a questionable methodology. They use a simple (Keynesian) multiplier of the expenditures undertaken by the cultural institutions themselves as well as of their visitors. Such an approach is at best able to capture the additional revenue created but not value added which should be the object of interest from an economic point of view. Impact studies also disregard the substitution effects produced. If, for example, a city organises a musical festival, it may well be that there is no overall effect on the economy as visitors may stop going to other festivals they attended before. Finally, impact studies disregard the positive and negative ex-

ternal effects produced by cultural activities. This is of major consequence because many "culturalists" argue that the critical effect of the arts is on people's preferences.

Another method to capture the influence of cultural institutions is to measure the social value created. Theatres, museums, and other cultural institutions provide benefits to persons beyond only those visiting cultural institutions. The literature distinguishes several such values:

– *Option value:* people value the possibility of enjoying art sometime in the future;
– *Existence value:* people benefit from knowing that a cultural supply exists but do not necessarily visit any cultural institutions now or in the future;
– *Bequest value:* people derive satisfaction from knowing that their descendents and other members of the future community will be able to enjoy the art supply if they choose to do so;
– *Prestige value:* people derive utility from knowing that the cultural supply in their city or region is highly valued by persons living outside. They themselves need not actually like the cultural offerings, nor even visit them;
– *Education value:* people are aware that the cultural supply contributes to their own or to other persons' sense of culture and value it because of that.

Cultural supply may also produce negative external effects, the costs of which are borne by persons not visiting the cultural venues. An example is the congestion and the noise pollution produced by the visitors to a museum, which is inflicted on the local community.

Most recently, the "Life Satisfaction Approach" (Frey, Luechinger and Stutzer 2008) based on the economics of happiness has allowed us to address the public good element of cultural supply. The advantage is that the benefits of culture are reflected in the independently measured life satisfaction indices, which are then related to the extent of cultural supply by econometrically estimating a happiness equation. In contrast, the willingness-to-pay-approach directly

* Professor for economics at the University of Zurich and research director of CREMA – Centre for Research in Economics, Management and the Arts, Switzerland.

links the benefits to a particular cultural supply. This procedure risks falling prey to the "prompting effect" identified in psychology: it may happen that as a result of questioning people in surveys a higher value is attributed to cultural supply than otherwise would be.

Analysing culture from the economic point of view

A *second approach* to relate the two terms "culture" and "economy" is to use the rational choice approach characterizing the economic approach. Cultural economics thus understood belongs to a more general field of the economic approach to human behaviour[1], comprising, for instance, the economics of education, health, the family, sports, or religion. The narrow formulation of the selfishly maximizing homo oeconomicus has often been extended in cultural economics by introducing psychological and sociological elements, including such features as decision anomalies, or new motivational elements such as the crowding out of intrinsic motivation by external interventions.[2]

Most adherents of "Cultural Economics" consider it a sub-discipline of economics[3] and combine the two approaches. In particular, they use the rational choice approach to analyze the effect of economic factors on the arts. Examples are the study of art markets, most prominently the rate of return on investment in art (paintings and other objects of art), the income of artists, or the impact of regulations on the arts.

The economics of art museums[4]

Demand for museums

There are two types of demand for museums. The first is the *private demand* exerted by the *visitors*. These may be persons interested in the exhibits as a leisure activity or as part of their profession as an art dealer or art historian. The visit may be undertaken by individual or family decisions, or may be part of an organised activity, e.g. schools or firms. The second type of demand comes from persons and organisations benefiting from a museum. This *social demand* is based on external effects and/or effects on economic activity.

Private demand

By far the largest number of museum visits can be attributed to leisure time activity. The number of visits can be analysed by a traditional *demand function*, capturing the major factors determining the rate of visits per time period. Its characteristics can be determined by maximising individual utility functions subject to budget and time constraints. Its features can be empirically measured by using the data on museum visits and the factors included in the demand function, normally by a multiple regression analysis. There are three major determinants relating to *prices* or *costs*:

(1) *Entrance fee.* Together with the number of visits, it determines the respective revenue gained. The price elasticity indicates by how many percent the number of visitors decreases when the entrance fee is raised by a given percentage. Econometric estimates for a large number of different museums in different countries suggest that the demand for museum services is price inelastic. The low price elasticity suggests that museums can generate significant increases in revenues through increasing admission fees.

(2) *Opportunity cost of time.* Visitors have to forgo alternatives when they visit a museum. In order to measure the monetary value, one must identify how much additional income could have been gained during that period. For persons with high income, potential and variable time use, mostly the self-employed, the opportunity costs of time are higher than for people of low income and fixed working hours. The latter are therefore expected to visit museums more often, all other things being equal. The opportunity cost of a museum visit not only depends on the time actually spent in a museum, but also on how much time is required to get to the museum, i.e. the location, the parking facilities, etc. For tourists, the opportunity costs of time tend to be lower than for local inhabitants, because they often visit a city with the purpose of visiting the respective museums.

(3) *Price of alternative activities.* These are, most importantly, substitute leisure activities, such as other cultural events (theatre, cinema), sports, dining out in restaurant, time spent with friends at home, etc. Even within the industry, museums may constitute a substitute for other museums. The higher the price of such alternatives is, the higher museum attendance is, cet. par. But complements also systematically influence the num-

---

[1] Pioneered by Becker (1976), see more recently Kirchgässner (2008), or Frey (1999).
[2] See Bénabou and Tirole (2006) and Frey (1997).
[3] See Baumol and Bowen (1966), Peacock (1993), Benhamou (2000), Throsby (2001), Frey and Pommerehne (1989), Frey (2004) and the collection of articles in Peacock (1998), Rizzo and Towse (2002), Towse (2003), Ginsburgh and Throsby (2006), Hutter and Throsby (2008).
[4] See more fully Frey and Meier (2006).

ber of museum visits. Important are the costs incurred through travel, accommodation and meals. The higher the costs are, the lower the rate of museum visits, cet. par.

Income is another "classical" determinant of the demand for museum visits. Econometric estimates reveal an income elastic demand, i.e., increasing real disposable income favours museums (see e.g., Withers 1980). There are many *other determinants* that may be included in a well-specified museum demand function, such as the attractiveness of the building, the level of amenities provided by a museum, i.e., the general atmosphere, the cafés and restaurants and the museum shop. Important are also the marketing efforts made by a museum, especially through regular and attention-catching advertising.

Social demand

Museums produce effects on people not actually visiting the museum. As already pointed out above, museums create *social values* (option, existence, bequest, prestige and educational values), for which they are not compensated in monetary terms. As a consequence, museums tend not to produce these values, or do so in too little quantity.

Museums may also produce negative external effects, whose costs are carried by other persons. An example would be the congestion and noise museum visitors inflict on a community.

Effects on markets

Museums produce *monetary values* for other economic actors. They create additional jobs and commercial revenue, particularly in the tourist and restaurant business. These expenditures create further expenditures (e.g., the restaurant owners spend more on food) and a multiplier effect results.

Supply of museums

The production of museum services shows some particularities.

Cost structure

Museums face a cost structure which differs from other firms in the service industry and can explain some of their particularities.

(i) Museums have high fixed costs and low variable costs. This leads to a diminishing average cost curve.

(ii) The marginal cost of a visitor is close to zero. Efficient pricing close to marginal cost therefore never covers the costs involved.

(iii) The costs of museums have a dynamic component, which is disadvantageous for the enterprises. Due to a productivity lag, museums, like most cultural organizations, face constantly increasing costs over time.

(iv) Opportunity costs constitute a substantial part of the costs of a museum. The exhibits of a museum generate high opportunity costs, but are seldom taken into account by the museum.

Museums own, through their collected art works, a huge endowment of high value. The works of art lead not only to storage and conservation costs, but also opportunity costs. The real costs of this capital stock would become apparent if museums borrowed money to buy the works of art. The annual interest, which the museum has to pay, constitutes the real costs of capital. The opportunity costs of a work of art, is its monetary value used in an alternative investment. The annual rate of return can be seen as the cost of the artwork. Other opportunity costs are, e.g., for the building and its alternative uses. For most museums, the value of their holdings is by far their greatest asset.

Most museums do not put a value on their collection in their accounts. Museums then understate their true capital costs (Grampp 1989, 171) by not taking opportunity costs into account. This practice leads to an understatement of the losses and an overstatement of potential revenues. It induces the museum to become too large.

Firm structure

Museums can take different organisational forms. Mainly, they can be private for-profit organisations, private non-profit organisations and public organisations run in a non-profit way. For Europe and for the United States, the non-profit organisational form is the predominant structure for museums.

Most museums face a demand curve lying below the average cost curve. This makes it impossible to set a price at which total admission receipts cover the total cost of the museum. If price discrimination is not applicable, or only of limited use, Hansmann (1981) argues that arts organisations can still ask individuals

for voluntary price discrimination. Visitors volunteer to pay more than the official admission price and thus become donors. The non-profit form dominates the for-profit enterprise in getting donations, because consumers lack exact information about the quality of the good and service provided.

Museum behaviour

The behaviour of a museum or its managers respectively can be modelled in two different ways:

(1) the neoclassical approach, which assumes rational actors maximising utility of a museum in a benevolent way; and
(2) an institutional approach, which goes beyond the market and emphasises the importance of institutional settings (e.g. the dependence on public support) for the behaviour of the museum management. We will present the first approach briefly and apply the second approach in more depth.

Neoclassical approach

The management of a museum is assumed to maximise a utility function. Assuming that a museum's objective is non-profit, the budget constraint requires zero net revenue. The non-profit structure of the museum raises the question as to what the museum manager maximises. The museum's utility is related to the number of visitors to the museum *(y)* and the quality of the exhibitions *(q)*. This assumes that the quality of the museum service can be measured. Then the decision by the museum management is to maximise

$$U = U(y,q)$$

subject to

$$p(y)y + g(q) + h(y) - c(y,q) = 0$$

The museum receives revenues from the entrance fees *(p)*, which is a function of the number of visitors *(y)*; the level of donations and government grants *(g)*, which depend exclusively on the quality of the museum; and the revenue from ancillary goods from the shop and the restaurant or café *(h)*, which depends on the number of visitors. Costs depend on both output and quality.

The first-order conditions can be written as:

$$U_y/\lambda + p_y y + p(y) + h_y = c_y$$

$$U_q/\lambda + g_q = c_q$$

$$p(y)y + g(q) + h(y) = c(y,q)$$

The subscripts indicate partial derivatives and λ is the multiplier on the constraint.

Two insights can be gained by looking at the optimality condition: Firstly, directors of a non-profit museum get extra utility from an increased number of visitors. They therefore set the entrance fee such that marginal revenue from entrance fees and ancillary goods are less than marginal costs. This result could explain why museums set too low a price according to the revenue maximising condition. Secondly, museums engage in increased quality beyond the point where marginal grant income is equal to the marginal cost of increasing the quality by one unit. This behaviour is due to the extra utility the museum gets from an increase in quality. According to the model, museums tend to provide too high quality at too low a price compared to revenue maximising firm.

The objectives of the museum, quality of the exhibition and number of visitors are the crucial assumption in the above model.

Institutional approach

Instead of taking for granted that managers of museums behave totally in the interests of the museums in the following model the management is primarily concerned with the personal utility of its members. The directors' utility depends on their own income and the prestige they receive within their reference group, which consists mainly of art lovers and the international museum community. A second source of amenity is derived from the agreeable working conditions and job security. But the museum management is not free to simply pursue its own goals, because they face certain constraints on their actions. Differences in these institutionally determined restrictions explain the museum management's behaviour. The finances available are the most important constraint on the museum's management. Other constraints, such as limited space or legal and administrative burdens imposed by the bureaucracy or labour unions, can also weigh heavily. The source of income differs considerably between museums. While some depend mostly on public grants, others rely exclusively on private money (donations and sponsorship, or income generated from entrance fees, shops and restaurants). From a politico-economic point of view, the institutional set up and the nature of funding of the museums has a dramatic influence on the behaviour of the management. We here distinguish three types of museums: public, private and museums dependent on donations.

The incentives for the museum's management to behave in a certain way vary enormously, depending on this institutional framework.

Public museums

Directors of purely *public museums* rely exclusively on public grants. The government allocates them sufficient funds to cover the expenses considered necessary for fulfilling their tasks. While they are expected to keep within the budget, if a deficit occurs, it will be covered by the public purse. This institutional setting provides little incentive to generate additional income and to keep costs at a minimum. The directorate will not allocate energy and resources generating additional income, because any additional money goes back into the national treasury. If they were to make a surplus, the public grants would correspondingly decrease, which acts like an implicit tax of 100 percent on profits. The museum's management tends to emphasise non-commercial aspects.

When the management is not forced to cover costs using its own efforts, it can legitimise its activities by referring to intrinsic "artistic", "scientific", or "non-commercial" standards. This helps the museum directors achieve their goal of gaining prestige, top performances and pleasant working conditions.

From this institutional point of view, one would therefore expect that:

– Public museums do not sell any paintings from their art collection because firstly, the directorate cannot use the income generated and secondly, activities are then measurable in monetary units, which leaves them open to criticism from the outside (be it by politicians or by public administrators).
– Directors of public museums are little interested in the number of visitors, because they are not dependent on income from entrance fees or shops. Therefore, exhibitions are designed to please an insider group of art "freaks".
– As a consequence, visitors' amenities in public museums are poorly developed. Little attention is paid to the profitability of museum shops, restaurants and cafeterias.

Private museums

Directors of purely *private museums*, on the other hand, have a strong incentive to increase their income, because their survival depends on sources of money like entrance fees, the restaurant, shop surpluses and additional money from sponsors and donors. If private museums generate a surplus, they are able to use it for future undertakings. As a result, it is to be expected that:

– Private museums rely on the market when managing their collection. Museums actively sell paintings that no longer fit into the collection and use the money to buy new works of art.
– Private museums are more concerned with attracting visitors. "Blockbuster" exhibitions guarantee that the museum will earn revenue, because the preferences of a larger group of people are taken into account. Hence, the exhibitions are better arranged from a didactic point of view, appealingly presented and, above all, the works of art are shown in a context, which is attractive to a large crowd.
– Private museums emphasise the visitors' amenities. The museum management is concerned with the well-being of the museum's visitors and tries to satisfy the preferences of the visitors at the lowest possible cost.

Museum dependent on donations

Contributions to non-profit museums may be deductible under the income tax rule for individuals and corporations in certain countries. When the marginal tax rate falls, the price for donations decreases, which reduces the willingness to donate. The tax-deductible status, if chosen by the museum, affects behaviour fundamentally. There is every incentive to avoid profits by charging low or "social" prices (which strengthens the legitimacy of tax-deductible status), while there is also an incentive to take out profits in the form of various kinds of excess payments that show up as costs.

Museum directors who depend on donations have an incentive to attract donors. People devote much effort and skilled resources to this end. Donors can be pleased in various ways, which influences the behaviour of the museum management. Donors can exercise some measure of control over the activities of museums.

– Donors directly influence museum policy in two ways: they can either interfere in the programming or they can set heavy legally binding limitations on the collections they donate. The limitations on the collections can have great impact on their man-

agement. Most donors want to highlight their own artistic visions. As the donations are partly financed by the government via their tax expenditures, the costs imposed by the donors on the museums are indeed a problem of supporting museums through tax deductions. Donors can be pleased when museums publicise the donor's contribution, thus enhancing their prestige. Museums have developed an elaborated system of honours ranging from appropriate attributes ("benefactor", "patron", "contributor", etc.), to naming rooms, wings and even whole buildings after the donor.

– Museums must give the impression that the donations are well used. Donors want to have the feeling that they contribute to a worthwhile cause. A good reputation of the art institution with the public and the media is crucial for the flow of donations. This forces the museum management to use their money efficiently. But there are no contracts completely controlling the directors. Donors therefore prefer to deal with non-profit firms acting under a "non-redistribution constraint" (i.e. prohibiting the personal appropriation of profits). Removing the profit goal avoids the problem of managers cheating the donors to some extent.

Conclusions

The intention of this contribution is to demonstrate that there is a well-developed and intellectually stimulating "Cultural Economics" able to provide new insights into the economic aspects of the arts. This approach has been illustrated through the example of the Economics of Museums.

References

Baumol, W. J. and W. G. Bowen (1966), *Performing Arts, the Economic Dilemma; a Study of Problems Common to Theatre, Opera, Music, and Dance,* Twentieth Century Fund; New York.

Becker, G. S. (1976), *The Economic Approach to Human Behavior,* University of Chicago Press, Chicago and London.

Bénabou, R. and J. Tirole (2006), "Incentives and Prosocial Behavior", *American Economic Review* 96(5): 1652–78.

Frey, B. S. (1997), *Not Just for the Money: An Economic Theory of Personal Motivation,* Edward Elgar, Cheltenham, UK.

Frey, B. S. (1999), *Economics as a Science of Human Behaviour: Towards a New Social Science Paradigm,* 2nd ed, Kluwer Academic, Dordrecht and London.

Frey, B. S. (2004), *Arts & Economics. Analysis & Cultural Policy,* 2nd ed., Springer, Berlin, Heidelberg, New York.

Frey, B. S., S. Luechinger and A. Stutzer (2008), "The Life Satisfaction Approach to Valuing Public Goods: The Case of Terrorism", Public Choice Online First.

Frey, B. S. and S. Meier (2006), The Economics of Museums, in V. A. Ginsburgh and D. Throsby, eds., *Handbook of the Economics of Art and Culture,* North Holland, Amsterdam, 1017–47.

Frey, B. S. and W. W. Pommerehne (1989), *Muses and Markets: Explorations in the Economics of the Arts,* Blackwell, Oxford.

Ginsburgh, V. and D. Throsby, eds. (2006), *Handbook of the Economics of Art and Culture,* North Holland, Amsterdam.

Grampp, W. D. (1989), *Pricing the Priceless: Art, Artists, and Economics,* Basic Books, New York.

Hansmann, H. (1981), "Nonprofit Enterprise in the Performing Arts", *Bell Journal of Economics* 12 (2), 341–61.

Hutter, M. and D. Throsby, eds (2008), *Beyond Price: Value in Culture, Economics, and the Arts,* Murphy Institute Studies in Political Economy, Cambridge University Press, Cambridge and New York.

Kirchgässner, G. (2008), *Homo Oeconomicus. The Economic Model of Behaviour and Its Application in Economics and Other Social Sciences,* Springer, Berlin, Heidelberg and New York.

Peacock, A. (1993), *Paying the Piper: Culture, Music and Money,* Edinburgh University Press, Edinburgh.

Peacock, A. (1998), *Does the Past Have a Future?: The Political Economy of Heritage,* Institute of Economic Affairs, London.

Rizzo, I. and R. Towse (2002), *The Political Economy of the Heritage: A Case Study of Sicily,* Edward Elgar, Cheltenham, UK, and Northampton, MA.

Throsby, D. (2001), *Economics and Culture,* Cambridge University Press, Cambridge; New York and Melbourne.

Towse, R. (2003), *A Handbook of Cultural Economics,* Edward Elgar, Cheltenham, UK, and Northampton, MA, US.

Withers, G. A. (1980), "Unbalanced Growth and the Demand for Performing Arts: An Econometric Analysis", *Southern Economic Journal* 46 (3), 735–42.

INTERNATIONAL EVIDENCE ON SCHOOL TRACKING: A REVIEW*

LUDGER WOESSMANN**

Introduction

Countries differ widely in the age at which they first track children into different types of schools. In the majority of OECD countries, tracking takes place at the age of 15 or 16, which means that there is no tracking until the 9th or 10th grade (Table). In contrast, some countries – including most parts of Germany – undertake the first tracking at the age of 10. Do such institutional differences make a difference for the level and distribution of educational outcomes?

This article reviews a series of comprehensive empirical studies that use the international variation to estimate the impact of tracking on student outcomes, as measured by various international student achievement tests.[1] As with other institutional features of the school system (see previous contributions to this journal (Woessmann 2003b, 2004)), institutional comparisons across countries lend themselves particularly well to analyze the effects of tracking.[2]

In this article, tracking refers to the placement of students into different school types, hierarchically structured by performance. Such school placement policies are variously called tracking, streaming, ability grouping, or selective (as opposed to comprehensive) schooling. The advantages and drawbacks of early tracking into differing-ability schools have been the subject of heated debates in many countries, and remain so today. It is all the more surprising how little empirical evidence is available about the actual effects of early tracking.

From a theoretical point of view, the effects of educational tracking are controversial. On the one hand, those who favour early tracking believe that homogeneous classes contribute to a focused curriculum and adequate progress, which could lead to an optimal learning situation for all students. On the other hand, it is conceivable that weaker students profit from the presence of better students without the latter suffering. Thus, critics of early tracking generally argue that the weaker groups are systematically disadvantaged if they are separated early on. Since theoretically, a wide variety of causal mechanisms is conceivable – depending, among others, on the type of peer effects assumed – the effects of educational tracking are ultimately an empirical question.[3]

In general, education systems aim both for a high level of achievement as well as equality of opportunity

** Ludger Woessmann is Professor of Economics at the University of Munich and heads the Department Human Capital and Innovation at the Ifo Institute for Economic Research.
[1] The tests investigated in this article are: PISA = Programme for International Student Assessment, TIMSS = Trends in International Mathematics and Science Study, and PIRLS = Progress in International Reading Literacy Study.
[2] See Woessmann (2003a, 2007b, 2007c) and West and Woessmann (2009) for empirically based analyses of additional reform possibilities in various areas of the school system.
[3] See Meier and Schuetz (2008) for a more comprehensive discussion of the theoretical background.

Table

School tracking in international comparison

10	11	12	13	14	15	16
Austria Germany	Czech Rep. Hungary Slovakia	Belgium Netherlands		Italy Korea	France Greece Ireland Japan Poland Portugal Switzerland	Australia Canada Denmark Finland Iceland Norway Spain Sweden United Kingdom United States

Note: Age when students are tracked for the first time.

Source: OECD (2004).

Figure 1

TEST SCORES AND ECONOMIC GROWTH

Notes: Relationship between average educational performance in numerous international student achievement tests and economic growth between 1960 and 2000, adjusted for the effects of the initial level of gross domestic product and years of schooling. Each dot represents a country.

Source: Based on Hanushek and Woessmann (2008).

for all children. That is why in this article, we investigate the effects of tracking both on the level and on the inequality of educational performance. To motivate these two goals, we start by briefly pointing out that both have important long-term consequences.

First, the average level of educational performance of the population is probably the most decisive factor for the long-term growth of an economy (see Hanushek and Woessmann 2008, 2009). Countries that had better results in the numerous international cognitive student achievement tests of the past – i.e., the predecessors of PISA – have a significantly higher long-term per-capita growth of their gross domestic product (Figure 1).

Second, the distribution of economic prosperity is closely related to the dispersion of educational performance. Thus, most of the international differences
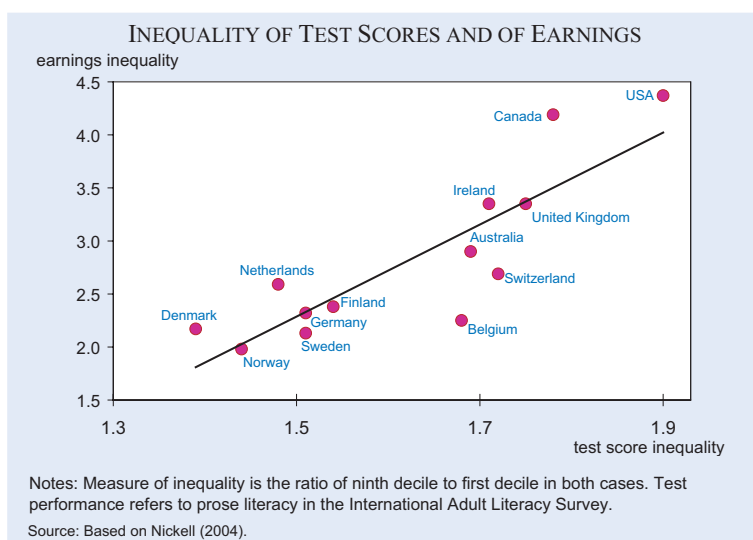
in earnings inequality can be attributed to international differences in the inequality of educational achievement as measured by test results (Nickell 2004). As shown in Figure 2, there is a close link between earnings inequality and inequality in educational performance. In contrast, minimum wages and the degree of unionisation among workers, for example, seem to have little to do with international differences in earnings inequality.

In order to investigate whether and how tracking influences educational performance and inequality, one needs variation in tracking. Since tracking is mostly a national feature, this article exploits the international variation depicted in the Table. In addition, the cross-country evidence will be compared to within-country evidence from a federal country, Germany, where tracking is a regional feature.

The next section analyzes in an international comparison whether early school tracking has an impact on how the level and distribution of student achievement at the end of lower-secondary school develops compared to the level and distribution already given at the end of primary school. The subsequent two sections summarize two studies that compare countries with and without early tracking in terms of whether tracking is associated with the dependency of students' achievement on their family background – i.e., whether tracking aggravates inequality of educational opportunity – once using the TIMSS studies, once using the PISA study. The final section investigates the same association across German states and compares the results with those of the international comparisons.

An international differences-in-differences approach

The first study reviewed here, by Hanushek and Woessmann (2006), places particular empha-

Figure 2

INEQUALITY OF TEST SCORES AND OF EARNINGS

Notes: Measure of inequality is the ratio of ninth decile to first decile in both cases. Test performance refers to prose literacy in the International Adult Literacy Survey.

Source: Based on Nickell (2004).

Figure 3

EDUCATIONAL INEQUALITY
IN PRIMARY AND SECONDARY SCHOOL
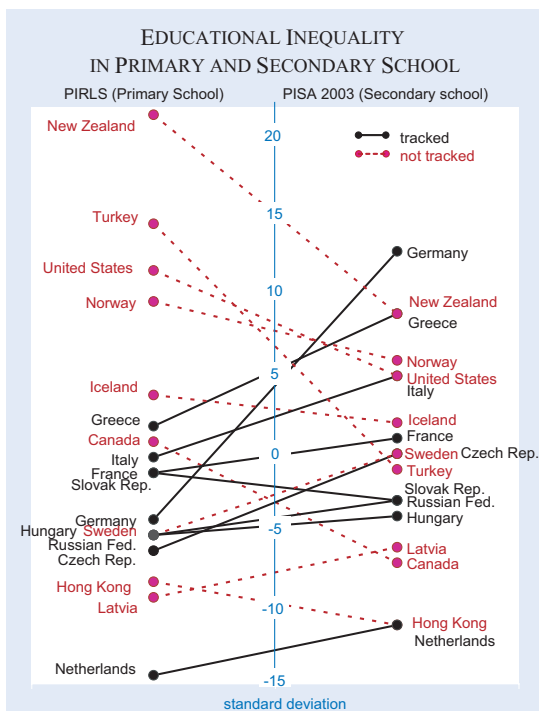
Notes: Dispersion of student achievement measured as standard deviation of test scores in primary school (PIRLS) in comparison to lower-secondary school (PISA 2003), in both cases measured as difference to the international mean of national standard deviations in each test. The lines indicate the change in performance dispersion from primary to lower-secondary school. The black solid lines indicate school systems that track their students into different school types before the age of 16, while the red dashed lines refer to those school systems that do not track their students by this age.

Source: Based on Hanushek and Woessmann (2006).

sis on the empirical identification of the causal effect of early tracking. They employ a method developed especially for estimating the effects of tracking to ensure that the relationship between early tracking and performance results in an international comparison is not caused by other differences in the educational systems of the countries investigated.[4]

The considerations of this approach start with the fact that in all countries, students are taught in a uniform school type for the first four years of schooling. It is only at a later point in time that tracking begins in some countries but not in others. Therefore, a comparison of the change in educational inequality between the end of primary school and towards the end of lower-secondary school can provide information on possible impacts of early tracking. The extent of inequality that exists in a country in 4th grade is clear-

ly not brought about by tracking.[5] As a consequence, it makes sense to take this general level of inequality out of the analysis and to consider only the change in inequality that occurs after 4th grade in order to determine the effect of early tracking on the extent of inequality in school performance. The question is thus how the dispersion of educational performance on international tests develops between the end of primary school and the end of lower-secondary school.

Basically, this method involves an investigation of the relationship depicted in Figure 3. The figure shows the inequality in educational performance in PIRLS and in PISA 2003 for all countries that participated in both of these studies. Both PIRLS and PISA investigate the reading competency of students. The measure for educational inequality within a country is the standard deviation in student test scores. It represents for every country the average deviation of the individual test results from the results of the national mean.[6]

The left-hand side of the figure depicts the inequality in the performance results of the PIRLS study, which measures reading performance of students in 4th grade. Note that at this point, Germany does not show any unusual inequality in students' performance. In 4th grade, Germany is still in the lower half of countries who have relatively balanced student performance. On the right-hand side, the figure shows the inequality in educational performance in the PISA 2003 study, which tests students at the age of 15, i.e., towards the end of lower-secondary school. As shown in the figure, Germany displays the largest inequality in the performance results at this stage.

The essence of the analysis now basically involves comparing the change in inequality that occurs from primary to lower-secondary school between countries with and without educational tracking during this period. When viewing the change between the performance dispersion in PIRLS and PISA, that part of the inequality measured at the end of lower-secondary school that already existed at the end of primary school is not taken into account. The change is indicated by the lines that connect the two points of each country. For countries with early tracking, black solid connecting lines are used, while red dashed lines indicate countries without early tracking. It is noticeable that nearly all the black solid lines point upwards

---

[4] For a more comprehensive report of this study, its methodology and additional results see Hanushek and Woessmann (2006).
[5] To minimize potential anticipation effects, we also performed the analysis with inequality in 3rd grade rather than 4th grade for a study that contains 3rd-grade results, obtaining similar qualitative results.

[6] Alternative measures of inequality, such as the difference in performance between the best and the worst 5 percent (or 25 percent) of students, yield similar results.

whereas nearly all the red dashed lines point downwards. This means that in countries that track their students into different school types, inequality increases systematically, whereas in countries where there is no tracking it decreases systematically (relative to the mean change in the sample). Of all countries, Germany is the one in which inequality increases the most between the end of primary school and the end of lower-secondary school.

Hanushek and Woessmann (2006) transform this graphic depiction into econometric estimates based on a differences-in-differences approach: The difference between countries with and without early tracking is investigated in terms of the difference in inequality between primary and lower-secondary school. The results show that early tracking leads to a systematic increase in inequality of student performance between the end of the primary and the end of lower-secondary school. This result is confirmed by various alternative model specifications and by the analysis of numerous additional pairs of international student achievement tests in primary and secondary school, including the various TIMSS tests. In total, eight different pairs of tests in primary and secondary schools were taken into account, combining a total of 176 observations.

In contrast to the results concerning the effects on inequality, the results on the performance level are less clear. But there is little evidence that early tracking would increase the performance level. To the contrary, in the most comprehensive model there is a negative effect of early tracking on the average performance level which is statistically marginally significant. Additionally, performance is evaluated at different percentiles of the performance distribution. The results indicate that not even for the best 5 percent of students is there a positive effect of early tracking.

The impact of family background on performance in TIMSS

In the previous investigation, the focus has been on the dispersion of student performance as a measure for the inequality of performance. Whereas in a positive analysis the results have a sizeable effect, it is not per se clear whether such a dispersion measure can be seen as normatively relevant because the dispersion does not necessarily tell us anything about the actual inequality of opportunity. That is why

Schuetz, Ursprung and Woessmann (2008) investigate a more direct measure for the inequality of opportunity: the extent to which individual student performance depends on the family background of the students.[7]

In national and international student achievement tests, it has been shown that almost without exception family background plays the most important role in influencing students' performance. The extent of family influence is, however, not unalterable. As reported below, there is considerable difference between countries. If the extent of family influence is seen as measure of inequality of opportunity that an educational system offers children from various family backgrounds, then countries vary in terms of the extent to which equality of opportunity is achieved. The study by Schuetz, Ursprung and Woessmann (2008) investigates whether these international differences in equality of opportunity are systematically related to the school policies of the countries.

The study combines the international TIMSS student achievement test of 1995 with the repeat study TIMSS-Repeat of 1999, which test performance in mathematics and natural sciences in 8th grade, at the age of approximately 14. Micro data are available for over 325,000 students from a total of 54 countries. The study uses the number of books available in a household as a proxy for socio-economic background. Since the multivariate analysis takes into account the effects of immigration status of the student, the mother and the father (as well as age, gender and family status), the estimates are limited to the association between student performance and family background *in the non-migrant population* of each country.

Figure 4 shows the resulting measure for inequality of opportunity in a country, measured as the effect of family background on students' achievements in mathematics and science for selected OECD countries.[8] The results indicate that in every country investigated, students with a high socio-economic background have a statistically significant lead vis-à-vis students with a low socio-economic background.

---

[7] Such an operationalisation draws on the concept of equality of opportunity as defined by Roemer (1998; see also Betts and Roemer 2007). The central idea of the concept is that inequality should only be tolerated to the extent that there are differences in an individual's effort – not, however, due to differences beyond the control of the individual (for example, as a consequence of gender, race, or family background). For details of the methods, data, and results of the reported study, see Schuetz, Ursprung and Woessmann (2008).
[8] For the estimated measure for inequality of opportunity in those of the 54 countries not shown in Figure 4, see Schuetz, Ursprung and Woessmann (2008).

Figure 4

EDUCATION POLICIES AND INEQUALITY OF OPPORTUNITY IN TIMSS

impact of family background

Notes: Estimated effect of three education policies on the inequality of educational opportunity (measured as the impact of family background on TIMSS student performance in test-score points), plotted against the entire extent of inequality of opportunity in selected OECD countries.

Source: Based on Schuetz, Ursprung and Woessmann (2008).

In addition, the results show clear differences in equality of opportunity between the countries studied. Germany, for example, ranks fourth among those OECD countries that are most unequal. Thus, Germany shows a particularly strong association between the performance of students and their particular socio-economic background, compared to other countries. The only two countries with greater inequality are the United Kingdom (England and Scotland) and Hungary. The countries with the greatest equality in this study are France and Canada.

How can we account for the fact that the equality of opportunity for children from different family backgrounds in the countries studied differs so widely? To analyze the possible effects of education policies on the differences in opportunity equality empirically, the study relates the measure of inequality of opportunity in the various countries to systemic characteristics of their education policies.

The study shows that inequality of opportunity is significantly smaller, the later the tracking age of students is. The sooner students are placed in different school types, the greater the dependence of student performance on family background.[9] If tracking is postponed by four years, for example, the impact of

family background on test scores declines by one quarter of the entire impact of the family background averaged across the OECD countries (see Figure 4). The international comparison also reveals that, apart from later tracking, a comprehensive system of pre-school education is another way of establishing larger equality of opportunity for children from different family backgrounds.

With this model, it is also possible to investigate the relationship between tracking and the average level of performance. The association between the age of the student when tracking begins and the average performance level of a country turns out not to be statistically significant. However, the point estimates are positive in the various models, indicating that the tendency is for the performance level to be lowered rather than raised by early tracking.

The impact of family background on performance in PISA 2003

Another study by Woessmann, Luedemann, Schuetz and West (2009) investigates the same association between tracking and equality of opportunity on the basis of micro data from the PISA 2003 study. In a comparable econometric specification, it investigates whether achievement in mathematics and science of the tested 15-year-old students is associated more closely with their family background if school tracking into different school types takes place at an earlier age. The analysis is based on data for around 180,000 students from 27 countries. The Index of Economic, Social and Cultural Status (ESCS) as provided by the PISA study is used as an alternative measure for family background.[10]

The qualitative results with the alternative data and measures are the same: the association between test scores and family background is significantly smaller, the higher the age of first tracking. This association is depicted in Figure 5: in countries with earlier track-

---

[9] Ammermüller (2005) reports similar results based on the international PIRLS and PISA data. Waldinger (2007) uses a combination of the approach of Hanushek and Woessmann (2006) and Schuetz, Ursprung and Woessmann (2008) and tends to find statistically insignificant results, but this may be largely due to limited degrees of freedom in samples of only 8–14 countries and a less informative tracking measure. Brunello and Checchi (2007) use the international approach described here for results beyond school age, finding that tracking increases the effect of family background on earnings in the labour market.

[10] See Woessmann, Luedemann, Schuetz and West (2009) for further details of this study.

Figure 5

TRACKING AND SOCIO-ECONOMIC STATUS IN PISA

performance in PISA 2003 scores
(relative to lowest category)

59.4

57.6

21.7

0.0

ten   age of first
        tracking

fifteen (at least)

low   high
socio-economic status

Notes: Low and high socio-economic status correspond to the 25th and 75th percentile of the PISA ESCS index, respectively. Performance difference in relationship to the lowest result category after exclusion of numerous additional factors.

Source: Based on Woessmann, Luedemann, Schuetz and West (2009).

ing, the difference in performance between children with different socio-economic backgrounds is considerably larger.

As the figure reveals, this effect arises primarily from the fact that children with low socio-economic status in countries with later tracking perform considerably better. At the same time, children from families with a relatively high socio-economic status perform at approximately the same level. Accordingly, with respect to the average performance level in PISA 2003, the overall effect of later tracking is positive, albeit not statistically significant (see Woessmann, Luedemann, Schuetz and West 2009).

A comparison of within-German and international evidence using PISA-E 2003

Can the cross-country analyses be trusted, or are there omitted factors at the country level such as differences in culture, language or legal background that may account for the reported associations? One way to test this is to perform analyses within a country with a common culture, language and jurisdiction in which the tracking regime differs across regions. Therefore, the study by Woessmann (2007a) investigates the association between school tracking and equality of opportunity in a comparison of German states *(Länder)*.[11] Even though tracking into different school types – usually *Hauptschule, Realschule* and *Gymnasium* – occurs at an early age everywhere in

Germany, there are small differences between the states. In Brandenburg and Berlin, there is a six-year comprehensive primary school, and only thereafter are the students separated into different school types. In all other states, tracking takes place already after 4th grade (in some cases with an orientation phase).

The advantage of an analysis within one country is that potentially remaining distortions of the international results due to unobserved, e.g., cultural, heterogeneity, can be excluded. Such a study is also particularly relevant for the German context. The disadvantage of such an analysis lies above all in the limited statistical degrees of freedom as a result of the small number of 16 state observations.[12]

In the state comparison, the slope of the so-called socio-economic gradient in the German PISA extension PISA-E 2003 is used as the measure for the inequality of opportunity. It indicates again how strongly student performance varies with the PISA-defined ESCS index of economic, social and cultural status. The steeper the slope, the stronger the influence of family background on test performance of students, and the smaller the extent to which the principle of equal opportunity is realized. To allow for a fair comparison between states, the socio-economic gradient is adjusted in a multivariate analysis for the facts that states with a heterogeneous population (such as city states) and those with a weak economy tend to achieve less equality of educational opportunities.

As the results reported in Figure 6 indicate, the adjusted socio-economic gradient in the two states with later tracking, namely Brandenburg and Berlin, is lower than in every other state. Thus, the association between student performance and socio-economic background is significantly lower in both states with later tracking. Also in Germany, it holds true that equality of opportunity for children from disadvantaged backgrounds is greater when school tracking occurs later.

---

[11] For details of the study reported in this section, its methodology and numerous further detailed results, see Woessmann (2007a).

[12] Unfortunately, the PISA-E micro data are not available in a form that allows a corresponding analysis to be made at the student level. Therefore, the 16 states constitute the level of observation of this study.

Figure 6

TRACKING AND INEQUALITY OF OPPORTUNITY
ACROSS GERMAN STATES

Note: Socio-economic gradients in the PISA-E 2003 mathematics test in states with tracking after grades 4 and 6, respectively, after excluding the effects of socio-economic background.

Source: Based on Woessmann (2007a).
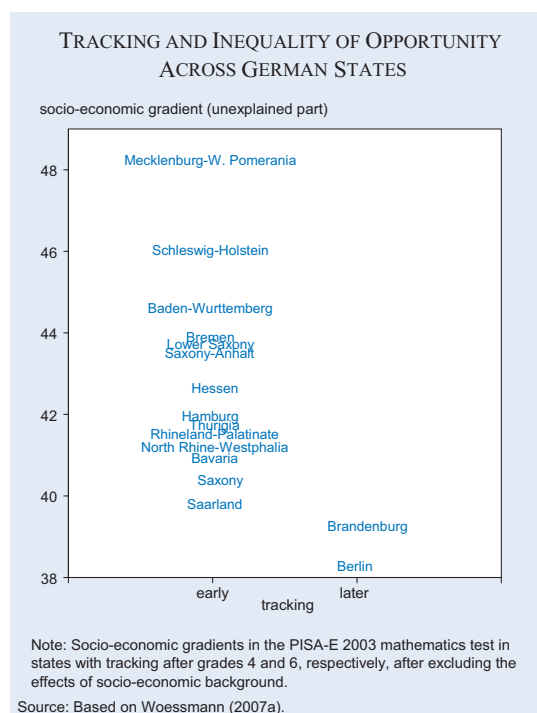
Similar to the international results, this does not come at the cost of lower average performance: There is no statistically significant association between the age of first tracking and the average performance level. The point estimate is negative, however, which again points more towards a performance-reducing than a performance-increasing effect of early tracking. The results for both equality of opportunity and performance level are extremely robust in terms of taking into account further control variables. Thus, consideration of indicators for city states or Eastern states and of other variables, ranging from the political nature of the state governments to the make-up of voters and the extent of expenditures for the school system, do not change the results.

Not only the age when school tracking sets in, but also the number of school types differs across German states. The Saarland, Saxony, Saxony-Anhalt and Thuringia, for example, have only two school types to choose from – i.e., there is only one other secondary school type in addition to the *Gymnasium*. As with the age when tracking begins, the German state study reveals a significant positive association of less tracking with equality of opportunity: In states with fewer school types, the opportunity for children from disadvantaged backgrounds is significantly higher, without the performance level suffering. Equality of opportunity is particularly lower when only a small share of students attends the lowest track, the *Hauptschule*.

In contrast, there is no systematic difference in equality of opportunity or performance level between states with and without the so-called *Gesamtschulen* (a "comprehensive" school type that is available in addition and next to the existing selective school types). Once further factors are taken into account, states with Gesamtschulen do not have a significantly lower performance. However, the traditional promise of putting *Gesamtschulen* next to the existing school types to improve equality of opportunity is not achieved either.

What is more, the study shows that the effects of early tracking and of the number of school types that are found in the comparison of German states are similarly significant in the international PISA comparison of the OECD countries. If the German states are analyzed together with the OECD countries in a sample of 42 to 54 observations, neither the effect of early tracking nor the effect of the number of school types differs significantly between the two observation groups. What is true for the international comparison is also true for Germany – suggesting that the international results are unlikely to be biased substantially by cross-country differences in culture, language, legal structures or the like.

Conclusions

The studies reviewed in this article investigate the international association of school tracking with the level and equality of opportunity of student performance. The results show that the dispersion of educational performance increases considerably more between primary school and lower-secondary school in systems that track students into differing-ability schools compared to school systems that remain comprehensive. In addition, the effect of family background on individual student performance is shown to be weaker, the later tracking into different school types takes place. Additionally, equality of opportunity tends to be higher in school systems that have fewer school types, once tracking has taken place.

At the same time, later tracking does not go hand in hand with a drop in the performance level. The estimated effects are generally statistically insignificant, but in all four studies they point towards better rather than poorer performance levels with later tracking. In general, however, tracked school systems appear to achieve approximately the same average level of student performance as non-tracked school systems.

The international evidence presented in this article is corroborated in various studies based on variations that exist within certain countries. In line with the within-German cross-state evidence presented here, Dustmann (2004) shows that in Germany, the choice of school type is closely associated with parental background, which translates into substantial earnings differences later in life. Bauer and Riphahn (2006) find that early tracking reduces intergenerational mobility in a comparison of Swiss cantons. Exploiting certain variations across schools within the Netherlands, van Elk, van der Steeg and Webbink (2009) find that early tracking reduces participation in and completion of higher education, without having a positive effect on those placed in the upper track.

Apart from the (usually limited) cross-sectional variation in some countries, another type of identification is possible if educational tracking varies over time within a country because of educational reforms. Thus, Meghir and Palme (2005) investigate the reform of educational tracking in Sweden in the 1950s and Pekkarinen, Uusitalo and Pekkala (2006) the Finnish reform in the 1970s. Exploiting the fact that both reforms were implemented gradually in different municipalities in a differences-in-differences framework, both studies find that the postponement of school tracking reduced later inequality on the labour market.[13]

Thus, both the international and several pieces of national evidence consistently show that the earlier the school tracking, the greater the impact of family background on educational success. The data show clearly that later school tracking increases equality of opportunity. At the same time, later tracking is not associated with a lower performance level. Thus, there is no indication of an apparent trade-off between the goals of equality and efficiency in the organisation of school systems.

---

[13] Another tracking reform that has been analyzed is the one in the United Kingdom (e.g., Galindo-Rueda and Vignoles 2007), but Pischke and Manning (2006) show that it is very hard to eliminate selection bias from different students attending different types of school in this kind of setting.

References

Ammermüller, A. (2005), "Educational Opportunities and the Role of Institutions", *ZEW Discussion Paper* No. 44, Centre for European Economic Research, Mannheim.

Bauer, P. and R.T. Riphahn (2006), "Timing of School Tracking as a Determinant of Intergenerational Transmission of Education", *Economics Letters* 91 (1), 90–97.

Betts, J. R. and J.E. Roemer (2007), "Equalizing Opportunity for Racial and Socioeconomic Groups in the United States through Educational-Finance Reform", in L. Woessmann and P. E. Peterson, eds., *Schools and the Equal Opportunity Problem,* MIT Press, Cambridge, Mass., 209–37.

Brunello, G. and D. Checchi (2007), "Does School Tracking Affect Equality of Opportunity?" New International Evidence, *Economic Policy* 22 (52), 781–861.

Dustmann, C. (2004), "Parental Background, Secondary School Track Choice, and Wages", *Oxford Economic Papers* 56 (2), 209–30.

Galindo-Rueda, F. and A. Vignoles (2007), "The Heterogeneous Effect of Selection in UK Secondary Schools", in L. Woessmann and P.E. Peterson, eds., *Schools and the Equal Opportunity Problem,* pp. 103–28, MIT Press, Cambridge, Mass.

Hanushek, E. A. and L. Woessmann (2006), "Does Early Tracking Affect Educational Inequality and Performance? Differences-in-Differences Evidence across Countries", *Economic Journal* 116 (510), C63–C76.

Hanushek, E. A. and L. Woessmann (2008), "The Role of Cognitive Skills in Economic Development", *Journal of Economic Literature* 46 (3), 607–68.

Hanushek, E. A. and L. Woessmann (2009), "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation", *NBER Working Paper* No. 14633, National Bureau for Economic Research, Cambridge, Mass.

Meghir, C. and M. Palme (2005), "Educational Reform, Ability, and Family Background", *American Economic Review* 95 (1), 414–24.

Meier, V. and G. Schuetz (2008), "The Economics of Tracking and Non-Tracking", *Zeitschrift für Betriebswirtschaft* 78 (1), 23–43.

Nickell, S. (2004), "Poverty and Worklessness in Britain", *Economic Journal* 114 (494), C1–C25.

Organisation for Economic Co-operation and Development (2004), *Learning for Tomorrow's World: First Results from PISA 2003,* Paris.

Pekkarinen, T., R. Uusitalo and S. Pekkala (2006), "Education Policy and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform", *IZA Discussion Paper* No. 2204, Institute for the Study of Labor, Bonn.

Pischke, J.-St. and A. Manning (2006), "Comprehensive versus Selective Schooling in England and Wales: What Do We Know?", *NBER Working Paper* No. 12176, National Bureau for Economic Research, Cambridge, Mass.

Roemer, J. E. (1998), *Equality of Opportunity,* Harvard University Press, Cambridge, Mass.

Schuetz, G., H.W. Ursprung and L. Woessmann (2008), "Education Policy and Equality of Opportunity", *Kyklos* 61 (2), 279–308.

van Elk, R., M. van der Steeg and D. Webbink (2009), "The Effect of Early Tracking on Participation in Higher Education", *CPB Document* No. 182. The Hague: Centraal Planbureau.

Waldinger, F. (2007), "Does Ability Tracking Exacerbate the Role of Family Background for Students' Test Scores?", Mimeo, London School of Economics.

West, M.R. and L. Woessmann (2009), "School Choice International: Higher Private School Share Boosts Test Scores", *Education Next* 9 (1), 54–61.

Woessmann, L. (2003a), "Schooling Resources, Educational Institutions, and Student Performance: The International Evidence", *Oxford Bulletin of Economics and Statistics* 65 (2), 117–70.

Woessmann, L. (2003b), "Central Exams as the 'Currency' of School Systems: International Evidence on the Complementarity of School Autonomy and Central Exams", *CESifo DICE Report – Journal for Institutional Comparisons* 1 (4), 46–56.

Woessmann, L. (2004), "Institutional Comparisons in Educational Production", *CESifo DICE Report – Journal for Institutional Comparisons* 2 (4), 3–6.

Woessmann, L. (2007a), "Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries", *Harvard University, Program on Education Policy and Governance Research Paper* PEPG 07-02, Cambridge, Mass. (available at www.hks.harvard.edu/pepg).

Woessmann, L. (2007b), "International Evidence on School Competition, Autonomy and Accountability: A Review", *Peabody Journal of Education* 82 (2–3), 473–97.

Woessmann, L. (2007c), "International Evidence on Expenditures and Class Size: A Review", *Brookings Papers on Education Policy* 2006/2007, 245–72.

Woessmann, L. (2009), "Beeinflusst Bildungsselektion Bildungsergebnisse und Ungleichheit? Internationale und nationale Evidenz", in M. Held, G. Kubon-Gilke and R. Sturn, eds., *Jahrbuch Normative und institutionelle Grundfragen der Ökonomik, Band 8: Bildungsökonomie in der Wissensgesellschaft,* Metropolis, Marburg pp. 147–167.

Woessmann, L., E. Luedemann, G. Schuetz and M. R. West (2009), *School Accountability, Autonomy and Choice around the World,* Edward Elgar, Cheltenham.

Regulatory Framework for Next-Generation Access Networks Across Europe*

Nina Czernich, Florian Engl, Oliver Falck, Thomas Kiessl and Tobias Kretschmer**

Introduction

Telecommunication networks around the globe are in transition from traditional public switched telephone networks to modern IP-based next-generation networks. This transition started in the backbone networks where increased demand led to a capacity shortage and the need for new fiber optic technologies. At present, fiber is standard in the backbone infrastructure, but the so-called last mile between the customer and the main distribution frame is generally still copper-based. The last mile is usually owned by the incumbent telecommunication carrier. However, due to the convergence of media and, eventually, increasing demand for high-speed Internet access, the last mile, too, has finally reached its physical limit and upgrading it to fiber is necessary. Eventually, the substitution of copper wire by fibre optic on the last mile, i.e., the construction of next-generation access networks, will complete the transition to the next-generation network.

This development raises the question of whether next-generation access networks should be regulated. The focus of the debate is whether competition

between telecommunication carriers encourages or discourages innovation and investment in new technology. On the one hand, it is argued that the incumbent's new infrastructure investment needs to be regulated, i.e., access should be granted to competitors so as to prevent the incumbent from again gaining a dominant position and concentrating market power in the telecommunications sector. On the other hand, it is claimed that investors need to be guaranteed pioneer rents as an inducement to make the investments in the first place and, thus, new investments should not be regulated.

This article compares different regulatory strategies for next-generation access networks across European countries (Germany, France, the Netherlands and the United Kingdom). In contrast to the United States, where next-generation facilities were deregulated in 2003, access regulation in Europe is still being discussed at both the national and European levels (cf. European Parliament 2008). Thus, this article will describe the different institutional settings with regard to access regulation of NGA across European countries.

We begin with a short technical description of the structure of the last mile. We then compare different regulatory strategies for the upgrade of the last mile to next-generation access networks and end with policy suggestions for regulation of telecommunication that will promote both competition and investment in next-generation access networks.

The structure of the last mile

The last mile is built hierarchically. Currently, many customers are connected via copper wire to a street cabinet; from there, the copper wires are bundled to the main distribution frame (MDF). At the MDF, the copper wires are connected to the fiber network via a digital subscriber line access multiplexer (DSLAM). The last mile is usually owned by the incumbent telecommunication carrier; however, the fiber backbone network may be owned by a competitor who, at the MDF, connects the customer, who has just "traveled" over the incumbent's last mile, to

Figure

THE STRUCTURE OF THE BROADBAND NETWORK

Source: Own Figure based on Anell et al. (2007).

its own fiber backbone network in a process known as local loop unbundling (LLU). Depending on the distance between the MDF and customer, a bandwidth of up to 16 MB/sec can be realized. The Figure summarizes the structure of the telecommunication network in transition to the next-generation network (NGN).

Bandwidth over the last mile can be increased by substituting fiber optics for the copper wire between the MDF and the street cabinet and installing a DSLAM in the street cabinet (fiber to the curb [FFTC]). This strategy makes possible a bandwidth up to 50 MB/sec (VDSL). Installing fiber optics for the entire distance between the MDF and the customer (fiber to the building [FTTB], or fiber to the home [FFTH]) removes any last-mile bandwidth restrictions (at least temporarily). With the completion of the next-generation access network (NGA), many components of the traditional access network become obsolete, including the MDFs, thus reducing the network's operating costs. This reorganization also affects competitors and their business models, of course, since they can no longer "pick up" their customers at the MDF.

How the incumbent telecommunication carrier arrives at NGA is heavily dependent on three factors: the structure of the traditional telecommunication network (i.e., the distance between MDFs and customers or the availability of alternative infrastructure), the demand for high-speed Internet and the regulatory framework for the NGA. Competitors with their own backbone network may either imitate the incumbent's strategy or engage in their own roll-

out strategy, a strategy often pursued by public utilities that use their own ducts and sewers to roll out a FTTB/FTTH network.

Roll-out strategies and regulatory framework for NGAs across europe

Germany

In Germany in 2007, there were 22.6 broadband subscribers per 100 inhabitants (Bundesnetzagentur 2009). With a share of nearly 92 percent of the market, DSL clearly is the dominant broadband access technology (Bundesnetzagentur 2009). Broadband access via cable has become more important since Deutsche Telekom sold its cable network in 2003. By the end of 2008, nearly 1.6 million cable broadband connections were in use (Bundesnetzagentur 2009). Among the DSL providers, Deutsche Telekom has a market share of about 51 percent, Deutsche Telekom's resellers have a market share of about 8 percent, and competitors with own infrastructure have a market share of about 37 percent (Bundesnetzagentur 2009). The latter usually have access to Deutsche Telekom's last copper mile at the MDFs. Additionally, some carriers (e.g., NetCologne, Wilhelm.tel) in large German cities provide FTTB. These carriers often use the municipal utilities' ducts and sewers to roll out their own fiber access network.

In 2005, Deutsche Telekom started to roll out fiber to the street cabinets in order to offer VDSL in large German cities. Where this has happened, the traditional street cabinets have been replaced by bigger cabinets that now accommodate the DSLAMs. Depending on the distance between the consumer and the street cabinet, this technology permits a bandwidth of up to 50 MB/sec. So far, the updated street cabinets are connected to the MDF by both fiber and the traditional copper wire. Copper wire is still currently used by Deutsche Telekom, its resellers and its competitors to offer access at a bandwidth of 16 MB/sec. Eventually, however, the copper wire will become obsolete and when its use between the street cabinet and the MDF is phased out, the very survival of Deutsche Telekom's competitors will be in doubt. Therefore, the German regulatory agency, Bundesnetzagentur, aims

to set incentives for more transparency regarding Deutsche Telekom's roll-out plans for the fiber access network (Bundesnetzagentur 2008).

The Bundesnetzagentur did not want to force Deutsche Telekom to allow competitors access to its new fiber access network. However, the European Commission filed an action against Germany, which, in turn, required Bundesnetzagentur to mandate subsidiary access (Möschel 2007; Cullen International 2007), that is, Deutsche Telekom must grant access to its ducts so that competitors can roll out their own fiber to the street cabinet within Deutsche Telekom's ducts. When there is insufficient space in the ducts to make this feasible, Deutsche Telekom is required to grant access to its own unused fiber (dark fiber). At the street cabinets, competitors have access to the copper pairs that connect the customers to the street cabinets. However, a competitor is required to build its own street cabinet and install a DSLAM if space will allow. Deutsche Telekom claims that the mandated subsidiary access to its new fiber network has made it hesitant to roll out additional fiber network.

France

At the end of the third quarter 2008, 17.1 million broadband connections existed in France, a rate of 26.9 connections per 100 inhabitants. 16.3 million of these connections were DSL connections; the remainder was split between cable, fiber, wireless and satellite (ARCEP 2008a). 8.2 million DSL connections are provided by competitors of the incumbent operator, France Télécom; 6.0 million of them by way of LLU and 2.2 million using bitstream[1] (ARCEP 2008b). At the end of 2007, DSL was available to 98.2 percent of the population, whereas cable was available to only 40 percent (ARCEP 2008c).

In its deployment of fiber, the French telecommunications market is one of the most competitive in Europe. Many different operators plan to roll out FTTH. This solution, which is more radical than the German step-by-step approach, is also favored by ARCEP, the French regulatory agency. The strategy of deploying FTTH instead of FTTC is based in the relatively long distance between the street cabinet and the consumer in most areas of France. In Germany, the average distance between the con-

sumer and the street cabinet is 300 meters; in France this distance averages 1 km. Therefore, roll out of fiber to the street cabinet would result in a far lower bandwidth in France than in Germany.

France Télécom began a pre-roll-out of FTTH in several large cities in 2007. Mass-market roll out is planned to take place from 2009 to 2012, aiming at 2 million subscribers by 2012 (cf. Elixmann et al. 2008). France Télécom's competitors are especially concentrated in the Paris region, profiting from Paris's unique and ubiquitous sewerage system that provides man-high canals to every house. This system makes the deployment of fiber very easy. Indeed, the carrier Free aims at covering 70 percent of Paris; Neuf Cegetel already reaches 400,000 homes in Paris and plans to reach 5 million by the end of 2012.

To spur investment in FTTH, ARCEP pursues three main strategies. First, France Télécom is required to grant access to its ducts. After surveying France Télécom's duct facilities in 10 cities, ARCEP concluded the ducts are sufficiently roomy that competitors can roll out their fiber in parallel. Using this space dramatically reduces competitors' costs of deploying fiber.

Second, ARCEP does not consider it economically reasonable to duplicate fiber roll out inside a building and thus desires competitors to coordinate on this front. This view is also held by the French parliament, which, in the Law on Modernizing the Economy (n° 2008–776), published 4 August 2008, mandated that new buildings must be equipped with fiber and that the first operator to deploy fiber inside a building must grant access on a non-discriminatory basis (ARCEP 2008d). By contrast, fiber unbundling is not mandated yet and there is no further obligation to provide bitstream access to the fiber network (Elixmann et al. 2008).

Third, municipalities are to play an active role in facilitating the fiber roll out by providing information about existing infrastructure and geographical characteristics, co-ordinating street work and negotiations with building owners, and granting access to their own infrastructure, such as sewer systems.

United Kingdom

The British telecommunication market includes 15.6 million broadband connections, of which 14.3 million were residential broadband connections

---

[1] In the case of bitstream access, the incumbent not only provides the unbundled local loop to the competitor but also transmission services. In comparison to LLU, bitstream access fosters a more service-based competition than facility-based competition.

at the end of 2007 (Ofcom 2008a). This results in a broadband penetration rate of 23.5 connections per 100 inhabitants in the United Kingdom. At the end of 2007, 26.3 percent of broadband connections used DSL from British Telecom; British Telecom's wholesale broadband products accounted for 26.9 percent of the market; 23.7 percent of broadband connections were provided by competitors using the incumbent's last mile; 21.8 percent used broadband over cable; and 1.3 percent had other types of connections (Ofcom 2008a). Due to increased DSL use (from 10 percent in 2006 to 23.7 percent in 2007), made possible by competitors using British Telecom's last mile via LLU, British Telecom suffered a massive decline in its wholesale broadband share, which had peaked at 42.3 percent in 2006 (Ofcom 2008a).

The emergence of facility-based competition based on LLU began in 2006 with the functional separation of British Telecom from its local access network. The local access network is now controlled by Openreach, a functionally separated branch of British Telecom. Openreach was founded to guarantee equal access to the local loop for British Telecom and its competitors. With the unbundling of 33.2 percent of all MDFs at the end of the first quarter 2008, competitors can now deliver LLU-DSL to 82.6 percent of all households (Ofcom 2008a).

Several market participants plan to roll out a NGA. In July 2008, British Telecom announced that it would invest EUR 1.78 billion in the roll out of FTTC for 7–10 million households by 2012 (Ofcom 2008b). FTTC allows a bandwidth of 40 Mbit/s at first. Newly built houses will be equipped with FTTH, which will allow a bandwidth of 100 Mbit/s and more. A first trial is taking place in Ebbsfleet Valley, Kent, where 10,000 new houses will be equipped with FTTH starting in 2008. In December 2007, Virgin Media, the British cable operator and British Telecom's main competitor, announced the upgrade of two-thirds of its cable network. This upgrade enables Virgin Media to offer a bandwidth up to 50 Mbit/s to 9 million households since the end of 2008 (Ofcom 2008a). In addition to British Telecom and Virgin Media, smaller entities, including H2O Networks Ltd and the Digital Region Project, are planning to roll out NGA networks in some areas. H2O Networks Ltd has plans to use the municipal sewerage system.

In a report published in September 2008, the British regulatory agency, Ofcom, elaborated on its vision of the regulatory regime during and after the roll out of NGA networks. Ofcom focuses on regulatory certainty, meaning that Ofcom's decisions will be "clear, timely and consistent over the longer term" (Ofcom 2008b). Ofcom acknowledges that standardized interfaces are very important in providing effective access opportunities for every market participant. Ofcom therefore relies on industry-driven development and supports the formation of industry groups, such as NGN UK or the Network Interoperability Consultative Committee (NICC), to harmonize and support the development of standards. In terms of access, Ofcom remains committed to promoting infrastructural competition at the deepest level possible. Where British Telecom deploys a FTTC network, Ofcom considers it obliged to provide access to the copper wire at the street cabinet. Where British Telecom deploys a FTTH network, Ofcom obligates it to allow fiber or wavelength unbundling or duct access, which means competitors are granted access to British Telecom's ducts so that they can duplicate the FTTH network for their own use. Ofcom is currently making a survey of British Telecom's ducts. In September 2008, the Caio Review (2008) recommended strategies to reduce the cost of deploying fiber, such as the co-ordination of street work and the provision of access to public utility ducts and sewers. Based on these recommendations, Ofcom is considering the use of municipal sewerage systems and is carefully examining the French approach. In addition to these "passive" access strategies, Ofcom is also considering "active" access strategies (e.g., bitstream access).

The pricing of access has yet to be decided. The prices of old access products were regulated by Ofcom, but the new prices will be unregulated, at least to begin with. Ofcom is aware of the risk of setting an inappropriate price and is also cognizant of the uncertainty and risk that the investor in NGA is exposed to and therefore wants the market parties to experiment with different prices. Excessive pricing does not seem likely due to regulation of other wholesale products, such as DSL access, and competition with cable and wireless NGN products.

Netherlands

The Dutch market for telecommunication is known for its well-developed broadband access. There are 33.5 broadband connections for every 100 inhabitants. Of the households that are connected to broadband, 60.8 percent chose broadband over DSL, 38 percent chose broadband over cable, and 1.2 percent chose a fiber connection (OPTA 2007). Compe-

tition with cable is especially fierce since 97 percent of Dutch households are covered by a cable-TV network (Kirsch and von Hirschhausen 2008). In 2007, the incumbent Dutch telecommunication carrier, KPN, dominated the market, with an overall broadband market share of 44 percent and a DSL market share greater than 70 percent (OPTA 2007). The remaining DSL providers, e.g., Orange or Tele2, run their own backbone networks and interconnect with KPN at the MDFs.

In November 2005, KPN declared that it will partly replace its copper access network with fiber optics by 2010.[2] According to this plan, fiber optics will either be rolled out to the street cabinets or to homes. This will give KPN a VDSL2 network that can deliver up to 50 MB/sec. As part of the upgrade, KPN intends to phase-out a great many of its 1,361 MDFs, at which competitors have unbundled access to KPN's network. However, the Dutch regulatory agency, OPTA, insists that a fully-fledged alternative be made available to competitors that will lose their interconnection at the MDFs. After considering the imposition of regulatory obligations on KPN, OPTA decided to stimulate a discussion among all market parties. In July 2007, KPN and the three largest MDF access operators (Orange, bbned and Tele2) reached a first memorandum of understanding that has since been accepted by 6 of the 10 market participants (KPN 2008).

Under the memorandum of understanding, KPN has agreed that it will not begin the phase-out of the MDFs currently used by competitors earlier than mid-2010. Afterward, the competitors can choose between the following various options. First, MDF access will continue at 196 MDFs, which will be transformed into so-called mini-MDFs. The cost of access shall remain at the same level as normal MDF access. Second, competitors can extend their own fiber optic networks to the street cabinets. If competitors choose this second option, KPN will guarantee a simultaneous roll out of its own fiber network and will bear the competitors' costs for migrating the unbundled lines from the MDF to the street cabinets. Additionally, KPN will pay a lump-sum compensation for the depreciation of the book value of MDF assets. A third option is that competitors return to providing bitstream access, which implies a step back in terms of

infrastructure competition, as competitors would have to purchase more services from the incumbent. Again, the costs for the migration to bitstream access will be borne by KPN, and KPN will also pay a compensation for the competitors' MDF assets.

OPTA acknowledges the efforts in the memorandum of understanding but argues that access at the street cabinet – even though it contributes to infrastructure competition – is not a good alternative for all competitors because it requires economies of scale (OPTA 2008a). Thus, OPTA will require KPN to guarantee bitstream access. In turn, KPN will be allowed to charge access fees that cover its costs and garner a reasonable profit (OPTA 2008a). In the areas where KPN is rolling out FTTH, access at the street cabinets will not be available. In these areas, OPTA wants KPN to grant unbundled access to the fiber optic network, which would be comparable to the old unbundling obligation of the copper cable at the MDF (OPTA 2008a). Further, OPTA is considering founding an industry body, similar to the NGN UK, in which all market participants meet to discuss the future standards of the NGN.

OPTA (2008b) has published a draft decision concerning the price regulation of access to unbundled fiber loops. The pricing regime is designed to protect against margin squeeze, price discrimination and excessive pricing. OPTA plans to impose a price cap calculated by using all-risk weighted average cost of capital. Under this scheme, in addition to the standard cost of capital, a premium based on the risk of the investment, as well as a premium regarding regulatory risk, will be considered (OPTA 2008b).

Conclusions

The discussion about regulation of NGAs across European countries basically revolves around three themes: access to NGAs, access to ducts and the phasing out of the traditional copper access network.

Mandating access to NGAs will discourage investment. As long as competitors will incur comparable costs for rolling out a fiber network, competition will ensue, even in the absence of access regulation. Against this background, granting regulatory holidays for NGAs for a given time period is a misplaced step; new infrastructure components along the NGA should be completely free of regulation. However, it is not economically reasonable to duplicate certain

---

[2] At the end of November 2008, KPN announced that it would take over 41 percent of Reggefiber, an operator of fiber networks in the Netherlands, and spend EUR 6–7 billion over the next five to seven years to roll out FTTH nationwide (cf. Dekker 2008). The joint venture of KPN and Reggefiber has yet to be approved by the Dutch competition commission, NMa.

parts of the NGA. This is most obviously true of fiber roll out inside a building, but also for some regions where duplication of a fiber access network would not be profitable. In these cases, national regulatory agencies should consider mandating competitor access to NGA at a regulated price. This price should include an adequate risk premium for the first-mover investor who bears the risk of uncertain future demand and should compensate for the fact that next-generation facilities can be rolled out at lower cost in the future. Another way of risk sharing is joint investment in NGA by all competitors, which would result in all competitors having access to this joint network at nondiscriminating prices.

Infrastructure competition will only occur if the costs to roll out the NGA are nearly equal for all competitors. Obviously, however, because the incumbent telecommunication carrier already has access to ducts, its costs will be lower than those of its competitors. Leveling the playing field will require symmetric access regulation of all existing ducts, including both those of the incumbent telecommunication carrier and alternative ducts or sewers owned, for example, by municipal utilities. If this symmetric access regulation to ducts guarantees that all competitors will face similar costs of rolling out a fiber network, different strategies for doing so will result, opening the possibility of innovation, instead of simple imitation of the incumbent's roll-out strategy.

The transition to NGA will make the traditional copper access network obsolete. This has serious implications as to the survival of competitors who at present rely on having access to the incumbent's local loop at the MDF. So that these competitors can make feasible plans about their future, it is important to ensure transparency as to the phasing out of the traditional copper access network. This could be achieved by sunset clauses that clearly set forth how long the copper access network will be available. As it is probably too optimistic to rely on a collective agreement between the incumbent telecommunication carrier and its competitors, national regulatory agencies may need to play a role in ensuring that appropriate transparency is achieved.

References

Anell, P., S. Jay and T. Plückebaum (2007), "Nachfrage nach Internetdiensten – Dienstearten, Verkehrseigenschaften und Quality of Service", *wik Diskussionsbeitrag* 302.

ARCEP (2008a), High Speed Internet Observatory – 3rd Quarter 2008 Provisional Results, http://www.arcep.fr/fileadmin/reprise/observatoire/haut-debit/trim03b-2008/obs-htdebit-t32008-eng.pdf (accessed 15 December 2008).

ARCEP (2008b), Report on DSL High Speed Wholesale Offers Unbundling and Bitstream 1st December 2008, http://www.arcep.fr/fileadmin/reprise/observatoire/blr/decembre-2008/tab-degroupage-011208-eng.pdf (accessed 15 December 2008).

ARCEP (2008c), ARCEP's *Annual Report 2007*, Paris.

ARCEP (2008d), Recommendations on the Implementation of Last Drop Sharing of the Last Part on Optical Fibre Networks, http://www.arcep.fr/uploads/tx_gspublication/recomd-mutual-ftth-1008-eng.pdf (accessed 15 December 2008).

Bundesnetzagentur (2008), *Jahresbericht 2007*, Bonn.

Bundesnetzagentur (2009), *Jahresbericht 2008*, Bonn.

Caio, F. (2008), *The Next Phase of Broadband UK: Action Now for Long Term Competitiveness,* London.

Cullen International (2007), *EU Telecommunications Flash Message* 75/2007.

Dekker, V. (2008), KPN en Reggefiber steken 6 miljard in national glasnet, http://www.trouw.nl/nieuws/economie/article1904419.ece/KPN_en_Reggefiber_steken_6_miljard_in_nationaal_glasnet_.html (accessed 15 December 2008).

Elixmann, D., D. Ilic, K.-H. Neumann and T. Plückebaum (2008), *The Economics of Next Generation Access – Final Report, Study for the European Competitive Telecommunication Association (ECTA),* WIK-Consult, Bad Honnef.

European Parliament (2008), *Report on the Proposal for a Directive of the European Parliament and of the Council Amending Directive 2002/21/EC on a Common Regulatory Framework for Electronic Communications Networks and Services,* Brussels.

Kirsch, F., and C. von Hirschhausen (2008), "Regulation of Next Generation Networks: Structural Separation, Access Regulation, or No Regulation at All?", Communications & Strategies 69, 63–85.

KPN (2008), *Annual Report and Form 20-F 2007,* Den Haag.

Möschel, W. (2007), "Fehlsame Weichenstellung in der TK-Regulierung", *Zeitschrift für Informations-, Telekommunikations- und Medienrecht* 9, 547–49.

Ofcom (2008a), *Communications Market Report,* London.

Ofcom (2008b), *Delivering Super-Fast Broadband in the UK,* London.

OPTA (2007), OPTA *Annual Report and Market Monitor 2007,* Den Haag.

OPTA (2008a), Fixed Telephony, Broadband and Leased Line – Preliminary Draft Decisions, http://www.opta.nl/download/Fixed+Telephony+Broadband+and+Leased+Lines+Preliminary+Draft+Decisions%5Fcontext+and+perspective+15+July+2008%5Fbeveiligd%2Epdf (accessed 15 December 2008).

OPTA (2008b), Draft Policy Rules—Tariff Regulation for Unbundled Fibre Access, http://www.opta.nl/download/draft%5Fpolicyrules%5Ftariffregulation%2Epdf (accessed 15 December 2008).

US Experience with Federal Budget Rules

Alan J. Auerbach[*]

Introduction

In its 2009 fiscal year, the United States will experience a federal budget deficit that could reach 12 percent of GDP or more,[1] an event unprecedented since the end of World War II. At the same time, long-term fiscal imbalances loom ever larger, as the ongoing demographic shift raises the old-age dependency ratio and medical expenditures continue to rise rapidly. Much of this year's deficit is attributable to the very severe recession, as a consequence of both automatic stabilizers and countercyclical discretionary fiscal policy actions already taken. But the US fiscal position had deteriorated even before the onset of recession at the end of 2007, as the federal budget surpluses of a decade ago gave way to substantial deficits that reached around 3.5 percent of GDP during the healthy fiscal years of 2003 and 2004.

What role have US budget rules played in this evolution of the fiscal picture, and what lies in store for the United States, as a new president and Congress confront the simultaneous challenges of recession and fiscal misalignment? A review of the experience of the United States under different budget regimes, and of the evolution of the regimes themselves, provides some clues as to how the situation may unfold.

Budget rules in the United States

Like the European Union under the Stability and Growth Pact (SGP), the United States has imposed fiscal rules in attempting to impose a degree of fiscal discipline on the political process of budget determination. But, in large part because of the differences in federal structure, the set-up in the United States differs from that in Europe. US budget rules are much more inflexible at lower levels of government than at the national level. Relatively tight rules apply in virtually all US states, where some sort of annual balanced budget requirement applies to current expenditures and incipient deficits must be dealt with either within the fiscal year or soon thereafter. These strong budget restrictions have induced strong state fiscal response to budget shocks.[2] Note that these state budget procedures are not imposed on the states by the US constitution or by other legislative mandates at the national level. Rather, they have been adopted individually by the states themselves and survived in the US political and economic environment that features substantial cross-state mobility and no explicit mechanism for policy co-ordination among the states. Thus, even though the state and local share of overall government activity has been rising over time in the United States, countercyclical fiscal policy remains within the purview of the federal government.

At the US federal level, there are no overriding provisions such as those of the SGP governing annual debt and deficits. Attempts over the years, particularly in the 1970s and 1980s, to pass a balanced-budget constitutional amendment never succeeded. Instead, the federal government has operated under a series of budget control regimes, typically featuring many components. Some components are of no apparent consequence, such as the federal debt limit (which must be raised when the government increases the amount of outstanding debt). Other components, discussed further below, may well have influenced fiscal policy choices. But none of the components have constitutional standing; they can easily be modified by the same government that is subject to them, and this is a main source of skepticism about their potential impact. US federal budget rules are also complex, owing to the separation of powers between the President

* Robert D. Burch Professor of Economics and Law and director, Robert D. Burch Center for Tax Policy and Public Finance, University of California, Berkeley.
[1] The Congressional Budget Office (CBO 2009) projects a deficit of 11.9 percent, including the large fiscal stimulus worked out in early 2009. US fiscal years begin on October 1 of the previous calendar year.

[2] See Bohn and Inman (1996), Poterba (1997), Auerbach (2003).

and Congress and also to the separate operating rules and procedures within Congress between the House of Representatives and the Senate. Thus, some space is necessary to characterize the key elements of the different budget regimes before discussing evidence of their possible effects.

Federal budget regimes

In 1974 Congress passed the Congressional Budget Act (CBA), introducing the first significant restraints on the US federal budget process. The CBA established a Budget Committee in each house of Congress and created the CBO to provide budget projections needed to implement the legislation. Under the CBA, both houses of Congress passed a resolution laying out limits on revenues and spending for the coming year, and subsequent legislation was supposed to adhere to these limits. The CBA provided a co-ordination mechanism for Congressional budget actions, and also introduced the practice of providing multi-year budget projections to Congress, a practice that eventually would play a role in the formulation of budget rules. However, the CBA did not restrict the size of government or the ability of government to increase spending or cut taxes. Indeed, by the 1980s, the United States was experiencing historically large (as of the time) peacetime budget deficits, with the deficit hitting 6 percent of GDP in 1983.

Concern about budget deficits led in late 1985 to the passage of the Gramm-Rudman-Hollings (GRH) bill, for the first time laying out specific deficit targets, which followed a declining path until a target of zero was called for in fiscal year 1991. The legislation required that the budget the President submitted each year be consistent with that year's deficit target and that Congress pass legislation in accord with the deficit target. If legislated policy was projected to miss the deficit target, then an automatic "sequestration" process would ensue, cutting the budget according to a specified allocation rule in order to meet the deficit target. The sequestration procedures were modified in 1987 after the first version of GRH was found unconstitutional by the Supreme Court. The 1987 legislation also relaxed the target deficit reduction path to one in which a zero deficit was to be achieved in 1993, rather than in 1991.

The idea behind the sequestration process was that it would represent so negative an outcome that the threat of its application would force Congress to

come to an agreement on legislation satisfying the deficit targets. In principle, sequestration was intended to have an important impact without ever actually being implemented, and the sequestration process never was initiated during the GRH period. However, as the revision of the trajectory of deficit targets in 1987 indicated, this approach to deficit control lacked an effective political mechanism for producing the desired results when the gap that needed to be closed was too big. Also, unlike the SGP, which provides some flexibility to accommodate the need to respond to economic downturns, GRH specified unconditional nominal deficit targets. So, as a recession began in the summer of 1990 and the tax system's automatic stabilizers caused revenue to decline, the difficulty of meeting GRH's ever-shrinking deficit targets forced a change in regime. In the fall of that year, a protracted "budget summit" meeting of President George H. W. Bush and leaders of Congress generated two important results: (1) a package of spending reductions and tax increases; and (2) the replacement of GRH with the Budget Enforcement Act (BEA), ushering in a new budget regime that was thought to overcome some of the problems of GRH.

The BEA eliminated annual deficit targets and instituted targets for discretionary spending, a category that excludes spending for health care, Social Security (public retirement and disability pensions), unemployment insurance and other "entitlement" programs. For the budget as a whole, BEA specified "pay-as-you-go" (PAYGO) restrictions on taxes and entitlement spending (other than Social Security), requiring that legislation on such items does not increase the deficit, in the aggregate. Except for discretionary spending, therefore, the budget rule now applied to legislated changes in policy, rather than to actual levels of spending or revenue. Changes in taxes or entitlement spending that resulted from economic growth, inflation, shifts in the income distribution or any other economic factors not directly attributable to policy actions were ignored when determining if the budget rules were satisfied. Thus, any cyclical or trend movements in the deficit, revenues or expenditures, except those associated with discretionary spending, were left outside the process. Given that discretionary spending by its nature involves little automatic response to the overall level of economic activity, this change meant that the need under GRH to offset automatic fiscal stabilizers was largely eliminated. On the other hand, the lack of an overall deficit target also meant that there was no "error-correction" mechanism to alter policy following an unstable

fiscal trajectory. Different underlying growth rates of taxes and spending, for example, faced no restrictions under BEA.

The BEA also introduced the use of a multi-year budget "window", requiring initially that the PAYGO requirement be satisfied over a five-year period, based on CBO projections, rather than just for the immediate fiscal year in which legislation was enacted. The aim was to incorporate the future effects of policy actions and to reduce the scope for using short-term timing changes to meet a one-year deficit target, a practice that had been encouraged by the structure of GRH. The BEA originally applied through 1995, but it was extended to 1998 and then to 2002 by legislation in 1993 and 1997, respectively, before officially expiring in 2002.

For much of the period during which the BEA was in force, the Senate followed additional rules with respect to its own operations, including a longer horizon (ten years) for the PAYGO rule and which made subject to a point of order (requiring a supermajority of 60 votes out of 100 to override) proposals that would increase the deficit beyond the budget window. This so-called "Byrd rule" became quite relevant in 2001, when the tax cuts proposed by President George W. Bush were adopted for only a ten-year period. Because of the possibility that Republicans would be unable to muster 60 votes to override the point of order based on the Byrd rule, the tax cuts were enacted to apply only during the budget window. Although there have been subsequent modifications to some provisions of the 2001 legislation, the law as of the beginning of 2009 still specifies that marginal income tax rates will rise in 2011 to their pre-2001 levels, and that the federal estate tax, having been fully phased out in 2010, will reappear in 2011 in its pre-2001 form. This case study provides compelling evidence of the impact of budget rules on the shape of legislation, although not necessarily on broader aggregates of spending, revenues and deficits.

Though officially in place through 2002, BEA began to erode after 1998, the fiscal year in which the United States had its first budget surplus since the 1960s. At first, the erosion took the form of procedures used to get around the BEA's restrictions, such as a large increase in 1999 in "emergency" discretionary spending that was not subject to the BEA caps on discretionary spending (CBO 1999). Eventually, however, Congress simply changed the budget rules as it went, adjusting the discretionary spending caps to conform to actual

spending and setting aside the PAYGO rules on a case-by-case basis. For example, Congress adopted the large tax cut proposed by President George W. Bush in 2001 without any offsetting revenue increases or entitlement spending reductions, even though these were required by BEA, which was still officially in force.

In the years immediately after 2002, Congress acted essentially without budget rules of the type embodied in GRH or BEA, even though it continued to use the annual budget plans the CBA requires to impose limits on the budget effects of certain legislation, as in 2003, when, in an episode discussed further below, a ten-year budget cost of USD 350 billion was imposed before the details of a tax cut were worked out. The 2007 change in the control of Congress led to a renewal of interest in budget rules, and more recent events, including President Obama's election and the severe recession and ballooning budget deficit, have left the potential future use of budget rules very much up in the air.

Can voluntary budget rules affect behavior?

The repeal of GRH in 1990 and the gradual collapse of BEA after 1998 both illustrate a characteristic of US federal budget rules: the rules cease to operate once they deviate too far from consensus policy. In 1990, GRH called for deficit reduction far greater than Congress wished to enact. After 1998, adherence to BEA would have resulted in significant budget discipline at odds with the politics of the time, given that the federal budget was in surplus and CBO was projecting even larger surpluses for the years to come. It is not surprising that the rules failed, given that they could be repealed by majority vote. The question is whether they had any significant impact at all.

It is possible, however, that a change in the budget process, even if adopted by simple majority, can change budget outcomes by altering each legislator's incentives. For example, suppose that each legislator prefers a low overall deficit to a higher one, but also wishes to promote his or her own spending priorities. With no budget rule in place, there may be no commitment mechanism in place to facilitate co-operation on keeping spending low. An overall spending limit could lead to an equilibrium outcome with proportionately lower spending and a low deficit, an outcome that legislators would prefer to the high-deficit-high-spending outcome with no budget rule. Thus, the outcome achieved under a budget rule might be consistent with

the contemporaneous wishes of the majority, while at the same time representing a different outcome than would occur without the budget rule in place.

The impact of US budget rules

Even if voluntary budget rules can have an effect by helping to sustain an equilibrium of fiscal responsibility, identifying the impact of such rules is made difficult by the rules' endogeneity. Changes in budget rules may be induced by the same factors that affect fiscal policy directly, so it is difficult to estimate the independent impact of the rules themselves without some independent source of variation in the rules, for which obvious candidates are lacking. With this obstacle in mind, my research has focused on more subtle types of behavioral responses, looking not at how overall deficits, revenues, and spending behaved during different budget regimes, but rather on the behavior of different components of revenues and spending, and on the responses of fiscal policy to macroeconomic and fiscal conditions. The argument is that the different budget rules should have predictable effects on these elements of behavior that simple changes in the degree of fiscal responsibility would not be expected to have. Put another way, there are a variety of coincidental effects of budget rules that one may view as exogenous, since they are distinct from the overall objectives of fiscal discipline.

As reviewed above, one can distinguish five periods: (1) the pre-CBA era, when no explicit budget rules applied; (2) the CBA period, from the 1974 CBA adoption until the 1985 passage of the first GRH Act, with co-ordinated budget policy but no exogenous restrictions on spending or taxes; (3) the GRH period, from late 1985 until the late-1990 adoption of the BEA, during which explicit one-year deficit targets were specified; (4) the BEA period, from adoption of BEA until its effective demise around 1999, during which discretionary spending caps and PAYGO rules for taxes and entitlement spending were in force; and (5) the post-BEA period from 1999 until recently, during which limited budget rules applied.

Non-defense discretionary spending

Non-defense discretionary spending is perhaps the most susceptible to budget restrictions, given that entitlement spending is not directly driven by annual appropriations, and defense spending depends very strongly on factors external to the budget process. Moreover, discretionary spending has figured differently in the various budget regimes identified. Absent any explicit budget rules, we might expect discretionary spending to increase with the health of the budget, as measured by the most recent budget surplus, if the surplus provides a signal of the resources available to the government. We might also expect discretionary spending to increase with the size of the output gap between potential and actual GDP, reflecting Keynesian objectives to stimulate the economy during periods of slow growth.

Table 1

Determinants of non-defense discretionary spending changes, 1963-2006

Dependent variable: Annual change in spending relative to potential GDP
(standard errors in parentheses)

Independent variable	Sample period					
	1963–2006	1963–1974 (Pre-CBA)	1975–1985 (CBA)	1987–1990 (GRH)	1992–1998 (BEA)	1999–2006 (post–BEA)
Constant	0.0011 (0.0005)	0.0036 (0.0007)	–0.0002 (0.0033)	0.0052 (0.0008)	–0.0012 (0.0014)	0.0010 (0.0003)
Budget surplus (–1)	0.0525 (0.0215)	0.1783 (0.0398)	–0.0199 (0.1431)	0.1605 (0.0232)	–0.0042 (0.1228)	0.0684 (0.0275)
GDP Gap (–1)	0.0278 (0.0171)	0.0668 (0.0165)	–0.0250 (0.0829)	0.0203 (0.0256)	0.0597 (0.1773)	0.0748 (0.0340)
$\overline{R}^2$	0.084	0.719	-0.229	0.953	0.155	0.377
Number of observations	44	12	11	4	7	8

Data Source: Congressional Budget Office.
Source: Auerbach (2008).

Table 1 provides estimates, based on annual fiscal year data,[3] of the impact of the prior year's budget surplus and output gap on the change in non-defense discretionary spending from the previous year, with all series scaled by potential GDP. The first column of the Table presents estimates for the full sample period, for which the expected relationship holds weakly. The relationship is much stronger for the period prior to the Congressional Budget Act, as illustrated in the next column of Table 1, and then disappears entirely during the CBA period. During the very short GRH period,[4] the very strong relationship of spending to the budget surplus reappears, but the response to the GDP gap does not. Although so short a sample period makes any conclusions tentative, both of these results are quite consistent with what one would expect, given the way that the GRH rules worked. As each year's budget surplus was required to hit a pre-specified target, any improvement in the condition of the budget made more resources available for spending increases or tax cuts. But an increase in the output gap had no such effect, because the deficit targets were not cyclically adjusted.

The behavior of discretionary spending under BEA was similar to that under CBA, not responsive to the budget surplus and not significantly responsive to the GDP gap. With discretionary spending caps in place, spending could only respond to the economy or to the budget if the caps themselves could respond, or if exceptions (such as emergency spending) could be arranged. The estimates suggest that neither of these channels was significant during the period, even though the caps were revised in 1993 and 1997 when the provisions of BEA were extended to later years. In the most recent period, after the effective demise of BEA, discretionary spending has reverted to a pattern of significant responses to both the budget surplus and the GDP gap. The response to the surplus is weaker than under GRH, but this makes sense, given that there is no explicit deficit target.

These results are consistent with budget rules having had some impact on non-defense discretionary spending. Except perhaps for the lack of responsiveness during the CBA period, the patterns are consistent with the restrictions imposed during the different budget periods, and so represent a more subtle form of evidence than that based on levels or the composition of spending, which might more easily be explained by alternative hypotheses.

Legislated changes in spending and revenue

Some elements of budget rules have involved levels of the deficit or its components. For example, GRH had deficit targets and BEA had caps on discretionary spending. But BEA also placed limits on legislated changes in spending and revenues, under its PAYGO rules. Thus, we should observe changes in patterns of these legislated changes if BEA had an impact. To construct measures of legislated changes in revenue and expenditure, I utilize data and procedures developed in earlier papers, including Auerbach (2008). CBO typically publishes two major revisions in its projections of revenue and spending each year, in late January or early February, and in August or September. Each revision indicates the changes from the previous forecast and divides these changes into components due to legislation and to other factors.

By accumulating changes attributed to legislative action between each of these forecasts, I derive continuous, roughly semiannual series of forecast revenue and spending policy changes. For each observation, I measure the policy change with respect to revenue and non-interest spending. As each update includes legislative changes for the current fiscal year and several subsequent years, these must be combined in some manner to provide a measure of the legislation's overall effect. I form the discounted sum of changes adopted during the interval for the current and subsequent four fiscal years (relative to each year's corresponding measure of potential GDP), with the five weights normalized to sum to 1 and a discount factor of 0.5.[5] Just as current policy changes have effects in future fiscal years, policy may respond to anticipated future fiscal conditions as well. Thus, as an alternative to the most recent budget surplus, I have found that a better fit results if one uses a measure based on the budget surpluses projected over the budget period, which are included in the CBO projections. To be consistent with the aggregate policy measure just developed, I aggregate the projected surplus for the current and next four fiscal years, as of the beginning of the period of observation, using the same discount factor as in the policy measure.

---

[3] All annual data are from the CBO.
[4] I exclude the fiscal year during which GRH was adopted (1986), as adoption was accompanied by a large spending cut that is difficult to attribute to the budget rule. I follow the same procedure below in excluding the fiscal year of the adoption of BEA (1991), which came out of a budget summit that also produced spending cuts and tax increases.

---

[5] That is, each successive future observation receives half the weight of the observation one period earlier. This discount factor was chosen in my earlier work based on goodness of fit.

Table 2

Determinants of policy changes, 1984-2007

Dependent variable: Semiannual policy change in revenue or non-interest spending relative to potential GDP
(standard errors in parentheses)

Independent variable	Sample period and dependent variable							
	1984:2–2007:2		1986:2–1990:2 (GRH)		1991:2–1999:1 (BEA)		1999:2–2007:2 (post–BEA)	
	Revenue	Spending	Revenue	Spending	Revenue	Spending	Revenue	Spending
Constant	–0.0012 (0.0003)	0.0021 (0.0005)	–0.0002 (0.0034)	0.0025 (0.0067)	–0.0010 (0.0007)	0.0006 (0.0005)	–0.0014 (0.0005)	0.0022 (0.0007)
GDP gap (–1)	–0.0700 (0.0210)	0.1198 (0.0308)	0.0659 (0.0755)	–0.0219 (0.1502)	–0.0860 (0.0504)	0.0501 (0.0381)	–0.1028 (0.0463)	0.1449 (0.0650)
Projected surplus	–0.0714 (0.0154)	0.1125 (0.0225)	–0.0297 (0.1142)	0.1126 (0.2272)	–0.0802 (0.0412)	0.0384 (0.0311)	–0.0857 (0.0407)	0.1193 (0.0571)
$\overline{R}^2$	0.298	0.335	0.242	-0.073	0.111	-0.018	0.164	0.164
Number of observations	47	47	9	9	16	16	17	17

Data Source: Congressional Budget Office.
Source: Auerbach (2008).

Table 2 presents results based on these constructed measures, starting with those for the full period, beginning with the observation for change in projections from winter to summer 1984, labeled 1984: 2, and ending with the changes in August, 2007. The explanatory variables are the beginning of period weighted projected surplus and the estimated GDP gap in the most recent quarter before the policy change being explained. The first column in Table 2 presents results with revenue as the dependent variable; the second column has the same specification but with non-interest spending as the dependent variable. Both columns show significant policy responses to both the budget surplus and the output gap, in the anticipated directions, with deficit-increasing policies resulting from higher projected surpluses or a higher output gap.

Data availability permits us to consider the performance of these equations for the three most recent budget regimes, GRH, BEA and post-BEA. The results for each of these regimes are presented in the remaining columns in Table 2.[6] For GRH, none of the coefficients are significant, but it is interesting to note that the coefficients on the GDP gap actually have the wrong sign, and do so only during this period.[7] As discussed above in relation to a similar finding for dis-

cretionary spending, with deficit targets not adjusted for the level of economic activity, there is no scope for countercyclical policy. Indeed, given that automatic stabilizers cause revenue to fall as output falls, the only way to keep the deficit from actually rising is to pass legislation to increase taxes or reduce spending as output falls – precisely the pro-cyclical legislative policy reactions estimated here.

Under the BEA regime, for which these particular data on legislative changes are perhaps the most relevant, significant impacts for both output and surplus variables are restored on both the revenue side and the spending side. This result appears at first to be somewhat puzzling. After all, if the PAYGO rules are in place, then how can changes in the projected budget surplus or the output gap have any net impact on legislated changes in the deficit? A potential answer to this puzzle is that the PAYGO restrictions did not apply directly to the revenue and spending variables being measured here.[8] First, the restrictions applied to legislation enacted in any given fiscal year, whereas the variables measured here are semiannual. Second, the PAYGO rules did not apply to the weighted sum of five years' revenue or spending changes but to unweighted sums over periods of different length. Thus, legislation was only partially restricted, so we might expect the overall response based on the five-year weighted average to be small-

---

[6] As before, I leave out the observations including the adoption of GRH (1986:1) and BEA (1991:1) to avoid attributing concurrent policy changes to the rules just being adopted.
[7] The results for GRH are similar when the lagged budget surplus, an arguably more relevant measure for this budget regime, is substituted in the equation for the weighted projected surplus.

[8] See Auerbach (2008) for further discussion.

er than with no restrictions but not zero. The last two columns of the Table are consistent with this conclusion, showing that all four policy responses strengthened after the demise of BEA. In summary, the strength and signs of legislative policy responses under different budget regimes are consistent with how the budget rules in each regime worked: pro-cyclical policy responses under GRH, full policy responses after BEA and muted policy responses under BEA.

Further effects on legislative behavior

The post-BEA period offers additional lessons concerning the effects of budget rule design. In particular, the long budget window used by the Senate during its post-BEA deliberations appears to have had an impact on the pattern of tax legislation, making so-called "sunset" provisions more common in tax legislation. Even though the PAYGO rule was no longer in force at the time, deliberations leading up to the 2003 tax cut included negotiations over the size of the tax cut and its components. An agreement was reached by the (Republican) leaders to limit the tax cut to a total revenue cost of USD 350 billion over the ten-year budget window, the total revenue cost being calculated using a simple sum over the ten years. This calculation method meant that there was a trade-off under the cap between the annual cost of the tax cut and the number of years over which the tax cut applied: a temporary tax cut could have a larger annual cost. Also, with no discounting of future revenue costs, tax cuts that applied only during the early years of the ten-year period were larger relative to the size of the economy than those that applied only later in the ten-year period. This lack of discounting, along with the greater uncertainty that future tax cuts could be sustained, made temporary tax cuts that applied early in the budget window more attractive to tax-cut proponents than tax cuts that were to be phased in only toward the end of the budget window. The 2003 outcome was a temporary tax cut expiring before the end of the budget window, illustrating that even weak budget procedures can affect the shape of legislation. One sees a similar impact looking at the GRH period: when only the immediate fiscal year was relevant to budget rules, deficit-reducing measures tended to be concentrated in that year (Auerbach 2008).

The objective in 2003 was to limit the size of the tax cut, not to encourage temporary policies. Likewise, the designers of GRH were interested in more than

temporary deficit reduction. In each instance, a multi-year budget window with discounting of future revenue costs might have led to a more rational outcome; it would have provided some credit for future years' deficit reduction under GRH, and would have reduced the cost of future tax cuts under the budget cap in 2003. Indeed, such a window is what is suggested by a model in which competing parties run deficits when in power to commit resources to their preferred objectives (Auerbach 2006), although the ideal parameters of such a mechanism would depend on a number of factors, such as the stability of government and the rigidity of existing policies.

What should budget rules accomplish?

It is difficult to determine the effects that US budget rules have had on aggregate spending, revenues, or deficits because of the endogeneity of the budget rules themselves as well as the many other political changes that occurred contemporaneously. The previous discussion does suggest, however, that the rules have exerted influences on fiscal behavior in ways consistent with what one would predict. In some cases, though, as with induced pro-cyclical policy under Gramm-Rudman-Hollings or the 2003 sunset provisions, there were some clear negative side-effects. These negative effects highlight that achievement of the underlying objectives of budget rules requires not only that the rules matter, but also that the rules are aligned with these underlying objectives.

Budget rules presumably are meant to relate to the growth of government and to the tendency to shift financial responsibilities to future generations. But none of the US budget rules studied here incorporates the implicit liabilities associated with the long-term commitments of entitlement programs. As a consequence, the entire period witnessed large increases in future implicit liabilities that had only limited impacts on short-term budget measures. Further, none of the measures have successfully solved the problem of how to incorporate a commitment to fiscal discipline with the need to maintain short-run flexibility to deal with macroeconomic conditions, a challenge that has dogged the SGP.[9] Both of these drawbacks of past budget regimes are now in the spotlight as the US fiscal policy confronts a recession and a long-run fiscal imbalance, each of great severity.

---

[9] In Auerbach (2006) I consider how this problem might be attacked in the context of a multi-year budget window by imposing shadow prices rather than absolute restrictions on annual deficits.

References

Auerbach, A. J. (2003), "Fiscal Policy, Past and Present", *Brookings Papers on Economic Activity* 34, 75–138.

Auerbach, A. J. (2006), "Budget Windows, Sunsets, and Fiscal Control", *Journal of Public Economics* 90, 87–100.

Auerbach, A. J. (2008), "Federal Budget Rules: The US Experience", *Swedish Economic Policy Review* 15, 57–82.

Bohn, H., and R. Inman (1996), "Constitutional Limits and Public Deficits: Evidence from the US States", *Carnegie-Rochester Conference Series on Public Policy* 45, 13–76.

Congressional Budget Office (1999), Emergency Spending Under the Budget Enforcement Act: An Update, US Government Printing Office, Washington.

Congressional Budget Office (2009), A Preliminary Analysis of the President's Budget and an Update of CBO's Budget and Economic Outlook.

Poterba, J.M. (1997), "Do Budget Rules Work?" in A. Auerbach, ed., *Fiscal Policy: Lessons from Economic Research,* MIT Press, Cambridge, MA.

Pension Reforms in OECD Countries

Since 1990, around half of OECD countries have undertaken far-reaching pension reforms. Most of them were packages comprising a number of different measures. Some of these changes, such as increases in pension ages, are highly visible and often politically controversial. Others, such as changes in either the way in which earnings are measured when calculating benefits or pensions are indexed, are more technical and less transparent. Some countries maintained the structure of the pension system, modifying only parameters and some of the rules, while others overhauled the entire system. The Table distinguishes between changes to parameters and changes to the paradigm of pension schemes.

Changing pension-system parameters

Changes in *pension age* are the most common feature of reform packages. The rationale for these changes is clear: starting in the 1960s, life expectancy began growing rapidly for both men and women, but many countries cut their retirement ages. Recent reforms have reversed the trend of lower pension eligibility ages, with ten countries introducing gradual increases in pension ages for both men and women. When these reforms are complete, most OECD countries will have a standard retirement age of 65 years, although in some countries the pension age is or will be 67 or more. Only France, Hungary and Slovakia plan to have normal pension ages below 65.

Nonetheless, effective retirement ages – the age at which people actually stop working – are lower on average than the standard pension age in most countries. A common policy response, adopted by nine countries, has been to encourage older workers to stay longer in their jobs by changing pension *incentives to retire*. Pathways to

early retirement, many of which were introduced in the 1970s in response to high and rising unemployment, have been closed to new entrants or restricted severely. Penalties for early retirement in old-age pension schemes have been introduced or increased in many countries. Some countries have increased the number of years of contributions required to receive a full pension. Other countries have introduced or increased the increments or bonuses paid to people retiring after the normal pension age (for more information see OECD 2007, Table II.1.1.).

Other changes to pension systems relate to the *calculation* of the earnings base for pension entitlements. Hungary based pension calculations on gross rather than net earnings. Japan extended pensionable earnings to include bonuses. Seven OECD countries have extended the period over which earnings are taken into account instead of just basing the benefit on a limited number of final-years or best salaries. Austria is gradually extending the averaging period from 15 to the 40 best years. France is moving from the best 10 years to the best 25 years. Finland, Poland, Portugal, Slovakia and Sweden are all moving to a lifetime average earnings measure. As a result of these reforms, most OECD countries will

Table

Main elements of pension reform packages in selected OECD countries

| | Changing parameters | | | | | | Changing paradigm | | |
| | Pension age | | Retirement incentives | Calculation | | Indexation | DC | NDC | Life expectancy |
	M	F		measure	revaluing				
Austria	•	•	•	•		•			
Finland			•	•	•	•			•
France			•	•	•				•
Germany	•	•	•						•
Hungary	•	•	•	•		•	•		
Italy	•	•	•			•		•	
Japan	•	•		•					
Korea	•	•							
Mexico							•		
New Zealand	•	•							
Poland			•	•		•	•	•	
Portugal		•	•	•	•	•			•
Slovakia	•	•		•			•		
Sweden				•			•	•	
Turkey	•	•							
United Kingdom	•	•	•						

Note: M = Male, F = Female, DC = defined contribution; NDC = notional defined contribution.

Sources: Martin and Whitehouse (2008), 8; OECD (2007), Table II.1.1.; Whiteford and Whitehouse (2006).

use a lifetime earnings measure. Furthermore, some systems revalue past earnings to take account of changes in living standards between the time pension rights accrued and when they are claimed. For example, France moved to price revaluation in the public scheme as early as 1985 and in the occupational schemes in 1996.

Finally, the way that pensions in payment are adjusted has been reformed. This process is called pre-retirement *indexation* but is also known as "valorisation". Many OECD countries have moved from adjusting pension benefits to earnings (earnings valorisation) towards full or partial indexation to prices (price valorisation). This preserves the purchasing power of pensions, but means that pensioners do not share to the same extent as workers in the general growth in living standards.

Changing pension-system paradigm

A number of countries opted for wholesale or systemic reform (Whitehouse 2007). The most common policy has been to remove all or part of the public defined-benefit (DB) pension system and replace it with *defined-contribution* (DC) provision. In DC schemes, the pension depends on contributions and the interest earned on them. Hungary, Mexico, Poland, Slovakia and Sweden have all introduced mandatory, privately managed individual accounts to replace part of the public pension.

Another change of retirement-income paradigm has been the shift in public pensions from DB plans to *notional accounts*. These schemes, adopted in Italy, Poland and Sweden, are designed to mimic some of the features of DC schemes. Hence, they are often called notional defined-contribution schemes (NDC). Again the pension depends on contributions but, unlike DC plans, the notional interest rate is set by government and often linked to wage or GDP growth. The schemes remain pay-as-you-go financed: no assets are accumulated.

The reforms of pension paradigms share one important feature: pensions will in future automatically adjust to changes in life expectancy. If life expectancy increases, the number of pensioners per contributor will increase and pension benefit will fall. When pension capital is accumulated in an individual account, it is usually transformed into a regular pension payment – an "annuity" – at retirement. An-

nuities will be lower, the higher life expectancy is at the time of retirement because the pension will be paid for a longer time. Benefits from notional accounts are calculated in a similar way. But such automatic adjustments to life expectancy can also be built into systems which have not undergone systemic reform. In Finland and Portugal, the value of pensions will be adjusted to changes in life expectancy at retirement. France, in the 2003 pension reform, linked the required number of years of contributions for a full pension with life expectancy. Germany will adjust benefits in a points system to reflect the financial sustainability of the pension system.

W.O.

References

Martin, J.P. and E. Whitehouse (2008), "Reforming Retirement-Income Systems: Lessons from the Recent Experiences of OECD Countries", *OECD Social, Employment and Migration Working Paper* 66.

OECD (2007), *Pensions at a Glance, Public Policies across OECD Countries,* Paris.

Whiteford, P. and E.R. Whitehouse (2006), "Pension Challenges and Pension Reforms in OECD Countries", *Oxford Review of Economic Policy* 22 (1), 78-94.

Whitehouse, E.R. (2007), "Life-Expectancy Risk and Pensions: Who Bears the Burden?", *OECD Social, Employment and Migration Working Paper* 60.

INTERNATIONAL MOBILITY OF THE HIGHLY SKILLED

In technological and economic development, human resources play a central role. Knowledge-based societies rely on highly-qualified labour forces. The growing intensity of knowledge means that all countries have a greater need for highly-skilled specialists who are able to access, understand and use knowledge. To make sure and improve the situation of countries in worldwide competition it is essential to support academic exchange and to attract highly-skilled specialists. Additionally, most OECD countries and other developed countries face a special challenge because of a higher need for the highly skilled in societies with shrinking populations. The OECD's *The Global Competition for Talent: Mobility of the Highly Skilled* analyses the mobility of the highly skilled and the policies of 15 countries that encourage the mobility of human resources in science and technology (HRST) and make themselves attractive for these groups.

Skilled HRST provide knowledge and contribute to innovative activity, thereby increasing economic growth and prosperity. Attracting more HRST can hasten the accumulation of knowledge, stimulate innovation and lead to higher levels of economic activity and prosperity. The loss of skilled people leads to concerns about labour force shortages and brain drain, particularly in developing countries.

Most OECD countries perceive mobility as important in terms of retaining and attracting HRST talent and have policies to assist and encourage mobility. Some make these policies part of an explicit strategy, often accompanied by a specific website (Australia, Austria, Belgium, Czech Republic, Finland, Nether-

Table 1

Mobility strategies

	Strategy to encourage mobility of HRST	Webpage or organisation providing information to inflows of HRST	Diaspora strategy
Australia	Initiatives in place to promote international research collaboration, including mobility of researchers.	Mobility portal established in conjunction with FEAST (Forum for European-Australian S&T Cooperation) – www.mobility.org.au	
Austria	Mobility mentioned in Programme of the Austrian Federal Government for the 23rd Legislative Period (2006-2008).	Yes. Primary site is the Researcher's Mobility Portal Austria (www.researchinaustria.at).	Networks for HRST in North America – ASciNA (Austrian Scientists & Scholars in North America) at www.ascina.at and OST (Office of Science and Technology) scientist network at www.ostina.org/content/view/7/26
Belgium	The relevant federated authorities put emphasis on their own priorities; however, mobility is a focus of policy efforts in each area.	Yes. Primary site is the Researcher's Mobility Portal Belgium (www.eracareers-belgium.be).	Networking events in the United States (Flemish government initiative), in the framework of the European ERA-Link project.
Canada	Mobility is central to a national strategy to make Canada one of the world's top countries for R&D and to build an innovative and competitive economy.		
Czech Republic	The need for mobility support is detailed in many recent documents *e.g.* Strategy of Economic Growth; National Reform Programme; National Innovation Policy for 2005-2010; National Development Plan of the Czech Republic.	Yes. Primary site is the Researcher's Mobility Portal Czech Republic (www.eracareers.cz).	
Finland	General goal of internationalisation in S&T and higher education.	Yes. Primary site is the Researcher's Mobility Portal Finland (accessed via Academy of Finland website www.aka.fi).	

(Table 1 continued)

	Strategy to encourage mobility of HRST	Webpage or organisation providing information to inflows of HRST	Diaspora strategy
Japan	Yes. "Strategic Promotion of the International Activity of Science and Technology" – Ministry of Education, Culture, Sports, Science and Technology (MEXT). High priority on strengthening collaboration with Asian countries and internationalising universities.	No central website.	
Korea	Enhancing international mobility of researchers in two categories: Korean natives who go abroad for study and research; and the inflow of foreign researchers.	No central website, but individual websites for each programme.	Web portal has been established for knowledge exchange among Korean researchers worldwide. Also distinguished scientists residing abroad are invited to visit Korea and establish networks with domestic researchers.
Netherlands	Ministry of Education, Culture and Research encourages mobility of researchers. Policy executed by Netherlands Organisation for Scientific Research (NWO).	Yes. Researcher's Mobility Portal: the Netherlands (www.eracareers.nl). Also NWO website (www.nwo.nl).	
New Zealand	Ministry of Research, Science and Technology currently developing a strategy to ensure New Zealand continues to attract, retain and develop top-performing people and teams in research, science and technology.		
Norway	No specific strategy, but the government assigns high priority to internationalisation of Norwegian research and invests large resources in international research programmes, in which mobility is a central measure.	Yes. Primary site is the Researcher's Mobility Portal Norway (www.eracareers.no).	
South Africa	Various mechanisms exist to encourage inward and outward mobility of researchers, doctoral students and HRST, mostly managed by the National Research Foundation (NRF).	NRF website (particularly the link to the South African Research Chairs Initiative) provides information on inflows of HRST (www.nrf.ac.za/sarchi/).	Platforms such as the African Union – African Diaspora Ministerial Conferences are used. Also, the Department of Science and Technology holds "South Africa Days" in several destinations abroad where there is South African talent.
Switzerland	No explicit strategy.	Yes. Portal for science, research and innovation in Switzerland (www.myscience.ch).	SwissTalents database of Swiss scientists and scientists with a strong link to Switzerland (www.swisstalents.org). Also www.myscience.ch.
United Kingdom	Yes. International mobility of students and researchers is embedded in "A Strategy for International Engagement in Research and Development", published by the Global Science and Innovation Forum (GSIF). See www.berr.gov.uk /files/file34726.pdf.	Yes. Primary site is Network UK – the researcher's mobility portal (www.britishcouncil.org/eumobility).	The GSIF Strategy for International Engagement in Research and Development recognises that the United Kingdom should encourage and promote an alumni network of researchers who have been working in the United Kingdom. A new International Fellowships Scheme, with linked alumni engagement and operated in partnership by the Royal Society, British Academy, Royal Academy of Engineering and Research Councils UK, will be launched in 2008/09.
European Commission	Yes. See European Commission COM(2001)331 (20/6/2001). Aim of strategy is to develop an open, trans-European labour market for researchers.	Yes. Primary site is the European Researcher's Mobility Portal.[a]	The European Commission has proposed to network all EU researchers working abroad, beginning with the United States (http://cordis.europa.eu/eralink/), followed by other countries, such as Japan.

[a] From June 2008, the principal website is the EURAXESS portal (http://ec.europa.eu/eracareers/index_en.cfm).

Source: OECD (2008).

lands, Norway, South Africa, Switzerland, United Kingdom, European Commission); others have no overall strategy. The members of the European Union appear to gain valuable leverage from EU initiatives, i.e., each country has a mobility portal that is linked to the wider European Researcher's Mobility Portal, which provides a wealth of information.

Some countries have specific strategies to maintain contact with their diaspora. They focus more on initial retention and re-attraction. Thus, in 2001 the Office of Science and Technology at the Austrian Embassy in Washington, D.C., set out to establish a network for Austrian expatriate researchers, the OST Scientist Network. The network serves as a foundation for a broader understanding of the needs of Austrian researchers in North America and provides a basis for recognising those needs in the Austrian scientific landscape. It advises on government-related matters and keeps members informed about science and technology policy issues. Additionally, it supplies news on new developments in Austrian and European research (see Table 1).

To support the inflow of the highly skilled, seven of the 15 analysed countries offer mobility opportuni-ties via their general high-skill migration policies. These are Australia, Canada, Czech Republic, Netherlands, New Zealand, Norway and Switzerland. In addition, almost all have targeted policies to assist HRST inflows, ranging from special visas to facilitated procedures that reduce delays or waive certain requirements. The Scientific Visa procedure adopted by European Commission Directive 2005/71 has been transposed into domestic legislation by Austria, Belgium, Czech Republic and the Netherlands (see Table 2).

Additionally, most of the countries provide economic incentives for inflows of HRST. The approaches vary widely, from a large number of policies across a number of policy categories to just a few programmes in selected categories. A special case is the Czech Republic, where no policies explicitly offer economic incentives to inflows of HRST. Australia's policy is focused on fellowships, while Finland and Switzerland have concentrated their efforts on only a few programmes that offer economic incentives. For Switzerland this is likely due to the attractive salaries and other conditions that are already successful in attracting researchers and other HRST to the country. Korea and New Zealand offer the broadest range of programmes: fellowships, grants,

Table 2

Immigration policy to facilitate inflows of HRST and economic incentives for inflows of HRST

	Facilitated procedures for HRST	Special visas for HRST	General highly-skilled migration policy	Fel-low-ships	Grants and project fund-ing	Scholar-ships and allow-ances	Tax bene-fits and subsi-dies	Other
				Number of programmes				
Australia		Yes – Business Long-Stay Visa, Educational Visa, Visiting Academic Visa	Yes	9				
Austria	Yes	Yes – Scientific Visa (under EC Directive 2005/71)		1	3			
Belgium		Yes – Scientific Visa (under EC Directive 2005/71)		3	3	2		1
Canada	Yes – Canada Chairs applicants exempt from Labour Market Opinion require-ment	Yes – Off-campus Work Permit	Yes	1	4			1
Czech Republic	Yes	Yes – Scientific Visa (under EC Directive 2005/71)	Yes					

(Table 2 continued)

	Facilitated procedures for HRST	Special visas for HRST	General highly-skilled migration policy	Fel-low-ships	Grants and project fund-ing	Scholar-ships and allow-ances	Tax bene-fits and subsi-dies	Other
				Number of programmes				
Finland	Informal arrangement based on co-operation between director-ate for immigra-tion and higher education institu-tions.				1		1	
Japan	Yes	Multiple-entry visa available		4	1	3		1
Korea		Yes – Science Card, IT Card, Gold Card		1	3	2	1	2
Netherlands		Yes – Scientific Visa (under EC Directive 2005/71)	Yes – a fast procedure for highly skilled with no work permit required. Facilitated procedures for family members.		4			
New Zealand	Yes – for occupa-tions on shortage lists		Yes	2	1	1	1	3
Norway	Facilitated proce-dures for accom-panying family members	Yes – work permit for researchers, scholarship holders and lecturers	Yes – skilled worker/specialist work permit quota scheme		3	3		
South Africa		2002 Immigration Act provides for work permits for foreign experts		1	2	1		2
Switzerland			Yes (for non EU-EFTA countries)		1			2
United Kingdom	Yes – no work permit require-ments for certain students	Sponsored Re-searcher work permit category for non-EEA nationals	Yes	4	3	5		6
European Commission		Scientific Visa (EC Directive 2005/71)			2			

Source: OECD (2008).

scholarships, tax benefits and other policies (see Table 2).

N.H.

Reference

OECD (2008), *The Global Competition for Talent: Mobility of the Highly Skilled,* Paris, pp. 122–144.

BANK SUPERVISION: MANNING OF THE SUPERVISORY AUTHORITY

In the ongoing debate on the causes of the current financial crisis, bank supervision plays an important role. There is consensus that bank supervision needs to be reformed. We briefly review the manning and activity of bank supervisory authorities around the world. The World Bank (2007) conducted a survey among supervisory authorities that depicts the situation at the end of 2006, thus just before the crisis started.

The manning of a supervisory authority plays an important role for its effectiveness. Here the big differences in the number of professional bank supervisors is of note (see Table), the absolute numbers of which shed light on their "power". Countries that experienced banking crises before seem to have more supervisors. An example is Sweden, which suffered from a banking crisis in the early 1990s and has 170 supervisors. France, a much larger country with many more banks to supervise, only has 160 supervisors. At the same time, some countries that attract a lot of foreign funds, such as Cyprus, Ireland and Luxembourg, have relatively few bank supervisors.

Onsite inspections are conducted in large and medium-sized banks either annually or every second year. Interestingly, the frequency of inspections does not vary systematically with the manning of the supervisory authorities. This suggests that the intensity with which inspections are conducted might depend on the personnel endowment of the supervisory authority.

Another relevant aspect is the experience of bank supervisors. There are two different measures of experience that complement each other. First, it is the number of bank supervisors that have more than 10 years of experience in bank supervision. Second, it is the average tenure of current supervisors. Both the share of experienced supervisors among all supervisors and their average tenure vary considerably. Especially in the new EU member states the share of experienced supervisors is relatively small, which is simply caused by the fact that their supervisory authorities themselves are rather young. In other countries it seems that bank supervision has expanded, which implies only relatively few experienced supervisors.

C.H.

Reference

World Bank (2007): Bank Regulation and Supervision Database, World Bank, Washington D.C.

Table

Bank supervision: Manning of supervisory authorities[a]

	Total number of professional bank supervisors	Frequency of onsite inspections conducted in large and medium size banks	Bank supervisors with more than 10 years of experience in bank supervision	Average tenure of current supervisors (i.e. average number of years current supervisors have been supervisors)
Austria	Not available	Every 2 years	Not available	Not available
Belgium	60	Annually	50	10
Czech Republic	About 100 employees of which 49 inspectors (32 onsite, 17 offsite).	Every 2 years	18 (of which 5 are onsite and 13 offsite inspectors).	6
Denmark	45	Less frequently	8	7.2
Finland	138 (total staff incl. securities markets supervisors).	Annually	66	10.3
France	160	Once per year	50	6
Germany	BaFin[b] 450; Bundesbank 1,003	Every 2 years	Not available	Not available
Greece	142	Every 2 years	100	20
Hungary	60	Every 2 years	8	7.1
Ireland	36	Annually	3	3.8

(Table continued)

	Total number of professional bank supervisors	Frequency of onsite inspections conducted in large and medium size banks	Bank super-visors with more than 10 years of experience in bank supervision	Average tenure of current supervisors (i.e. average num-ber of years current supervisors have been supervisors)
Italy	Banking supervision is carried out by 3 departments of the Bank of Italy's central ad-ministration and by local branches (one in each Italian province). There are more than 400 bank supervisors, excluding the branches.	Less frequently	Not available	Not available
Luxembourg	30	Annually	17	10.7
Netherlands	92	Annually	18	6
Poland	464	Every 2 years	279	10.3
Portugal	Total 171 and related directly to supervisory activities 72.	Every 2 years	Total 121 and re-lated directly to supervisory activi-ties 46.	16
Slovakia	85	Every 2 years	10	5
Spain	300	Not available	150	10
Sweden	170	Annually	80	Not available
United King-dom	Not available	Every 2 years	Not available	Not available
Norway	Not available	Annually	60% of the onsite inspectors	7
Switzerland	132	Annually	45	10
Australia	248	Annually	Not available	5.9
Canada	Ca. 210 (450 total FTE at Office of the Superintendent of Financial Institutions)	Annually	11	7
Japan	450	Annually; As an approx-imation, for large banks: Annually, for medium size banks: every 2 years.	Not available	Not available
New Zealand	7	Less frequently	4	10
United States	2,218	Annually	1,269 (OCC[c]) and 787 (Fed[d])	16.1 (OCC[c]) and 8.8 (Fed[d])

[a] The data from the survey was available in early July 2007 and it is perhaps accurate to interpret the responses as de-scribing the situation as of 2006. – [b] Bundesanstalt für Finanzdienstleistungsaufsicht. – [c] OCC: Office of the Comp-troller of the Currency. – [d] Fed: Federal Reserve System.

Source: World Bank (2007).

BANK SUPERVISION: AUTHORITY AND LIABILITY OF BANK SUPERVISORS

The current financial crisis has touched off a debate about the future international financial architecture. An integral part of future reforms is the design of bank supervision, which will take place on many layers. Here we present several features of the organisational design of bank supervisory authorities (see Table). These features were compiled by the World Bank (2007) and best capture the state of affairs at the end of 2006, just before the current crisis started.

Bank supervisors ensure that banks fulfil the requirements of prudential regulation that aim at limiting the level of risk bank creditors are exposed to. Supervisors must report infractions of regulations that are found in the course of supervision. In about half of the countries considered there are mandatory actions that must be taken by the supervisor in these cases. In other countries the supervisors have more leeway in their reaction to infractions.

Another important feature is the liability of both the supervisory authority itself and the individual supervisory staff for damages to a bank caused by their actions or omissions committed in the good faith exercise of their duties. There are only a few countries in which the individual supervisor can be held personally liable in these cases. In many more countries the supervisory authority is liable in such cases. Here, the legal origin of a country seems to matter because in several continental European countries the supervisory authority can be held liable whereas in the Anglo-Saxon countries this is not the case. Liability rules have important effects on the incentives of the individual supervisor or

Table

Supervision: Authority and liability of bank supervisors[a]

	If an infraction of any prudential regulation is found in the course of supervision, must it be reported?	Are there mandatory actions that the supervisor must take in these cases?	Can individual supervisory staff be held personally liable for damages to a bank caused by their actions or omissions committed when exercising their duties in good faith?	Can the supervisory agency be held liable for damages to a bank caused by its actions?
Austria	Not available	Not available	No	No
Belgium	Yes	No	No	Yes
Czech Republic	Yes	No	Yes	Yes
Denmark	Yes	Yes	No	Yes
Finland	Yes	Yes	No	Yes
France	Not available	No	No	Yes
Germany	Yes	No	No	Yes
Greece	Not available	Yes	Yes	Yes
Hungary	Yes	No	No	Yes
Ireland	Yes	Yes	No	No
Italy	Yes	Yes	Yes	Yes
Luxembourg	Yes	Yes	No	Yes
Netherlands	Yes	Yes	No	No
Poland	Yes	Yes	No	Yes
Portugal	Yes	No	No	Yes
Slovak Republic	Yes	Yes	No	Yes
Spain	Yes	Yes	No	Yes
Sweden	Yes	No	No	Not available
United Kingdom	Yes	No	No	No
Norway	Yes	No	No	Yes
Switzerland	Yes	Yes	No	No
Australia	Yes	No	No	No
Canada	Yes	No	No	No
Japan	Not available	Not available	No	Not available
New Zealand	Yes	No	No	No
United States	Yes	Yes	No	No

[a] The data from the survey was available in early July 2007 and it is perhaps accurate to interpret the responses as describing the situation as of 2006.

Source: World Bank (2007).

the supervisory body as a whole. Therefore they should be carefully considered when drafting new rules for bank supervision.

C.H.

Reference

World Bank (2007): Bank Regulation and Supervision Database, World Bank, Washington D.C.

ON EUROPE'S GAS (IN-)SECURITY

The recent gas crisis, with its attendant feeling of déjà vu, once again laid bare both the extent to which Europe is dependent on Russian gas imports as well as its lack of and need for integrated risk management mechanisms and proper national and supranational regulation. Natural gas covers a significant portion of Europe's energy needs (25 percent on average in the EU-27, with up to 44 percent in some countries) most of it from non-European sources, foremost from Russia. And the volume of gas imports to Europe is increasing, as well as the distances over which it must be transported. At the same time domestic gas production, and the flexibility that it offers, is decreasing. All this makes Europe highly susceptible to supply disruptions. Reserve storage capacities, an important instrument to offset such disruptions, differ markedly from country to

country. And such reserves can make all the difference: during the latest supply crisis, some countries had reserves to meet their needs for several months, while others did not have enough even to last a single week (see Table).

All of this underscores just how badly Europe needs integrated risk management mechanisms and proper national and supranational regulation.

Although Europe cannot be expected to resolve its energy dependence problem in the short run, there still are measures that can be taken to dampen supply shocks. Increasing and better integrating European storage capacities would be an important step in the right direction. Some countries apparently coped with the disruption in gas supply much better than others. While Germany announced that it could run half a year on its reserves, others had problems ranging from the merely severe to veritable national emergencies. This is undoubtedly due to differences in resource endowment as well as to the

Table

Natural gas storage capacities of European member states

	Indigenous production (% of TNS*)	Natural gas consumption (% of TPEC**)	Non-EU imports (% of TNS***)	Natural gas consumption per day (mcm ****)	Number of storage facilities	Max. working volume (mcm)	Days covered by storage volume	Max. withdrawal capacity (mcm per day)
Austria	22	23	109	24	4	2,820	119	33
Belgium	0	26	69	44	2	779	18	46
Bulgaria	15	14	85	9	1	1,000	117	8
Czech Republic	1	19	105	25	9	3,376	132	55
Denmark	204	22	0	14	2	881	63	22
Estonia	0	13	100	2	0	0	0	0
Finland	0	11	100	12	0	0	0	0
France	2	14	84	120	15	11,700	98	200
Germany	18	23	75	239	44	19,138	80	463
Greece	0	7	100	7	1	75	11	5
Hungary	21	42	74	36	5	3,400	95	48
Ireland	10	29	0	14	1	198	15	3
Italy	13	36	80	212	10	13,400	63	152
Latvia	0	32	100	5	1	2,325	486	25
Lithuania	0	29	100	7	0	0	0	0
Luxembourg	0	29	0	4	0	0	0	0
Netherlands	161	44	27	104	3	3,500	34	30
Poland	31	13	70	37	7	1,651	44	34
Portugal	0	16	110	14	1	90	7	7
Romania	71	35	30	30	6	2,850	95	40
Slovakia	1	31	104	18	3	2,066	114	32
Slovenia	0	14	101	3	0	0	0	0
Spain	0	21	105	91	2	1,659	18	12
Sweden	0	2	0	3	1	10	4	1
United Kingdom	89	39	22	277	9	4,364	16	127

* TNS: Total Net Supplies. – ** TPEC: Total Primary Energy Consumption. – *** Can exceed 100% because total net supplies include changes of stock. – **** mcm: million cubic metres.

Source: Own compilation on the basis of data from Eurogas's *Annual Report 2006–2007*, www.eurogas.org, accessed 22 January 09.

availability of, or lack of, diversity of suppliers, but also to huge differences in storage capacities: Germany, Italy and France possess significant installed capacities; Portugal, Sweden and Greece have fairly meagre ones, and Estonia, Finland, Lithuania, Luxembourg and Slovenia have no storage capacity at all.

The first oil crisis in the 1970s prompted industrialised countries to create a crisis management outfit – the International Energy Agency – that is supposed to initiate actions to cope with emergency situations. The IEA member countries are obliged to hold mandatory oil stock levels equivalent to at least 90 days of their net imports and to have a clear mechanism for collective actions in case of emergency.[1] Why are there no storage rules for the gas sector? Gas storage is – for physical reasons – much more expensive than oil storage. According to the IEA, the capital cost of gas storage is between five to seven times the cost of underground oil storage facilities per tonne of oil equivalent (toe) stored. Liquefied natural gas (LNG) storage is even more expensive: the capital cost of LNG storage is ten times higher than the cost of storing oil in tanks and approximately fifty times the cost of underground oil storage per toe stored. Besides the sizable fixed cost of gas storage, there is also a relatively high variable cost that comes on top of it.[2]

Despite the high cost, investment in gas storage can actually be highly profitable, as seems to be the case in North America, where storage capacity is expanding fast. In most of continental Europe, in contrast, storage investment is lagging substantially. This is largely due to Europe's suboptimal market structure. Its national gas markets are in the midst of a transition from the old mono/oligopolistic structure to a North-American-style competitive one. While some European countries like the UK, Belgium, the Netherlands, and Spain have made good progress in

terms of liberalisation, most of Europe still remains dominated by long-term take-or-pay contracts, with prices adjusted periodically on the basis of a linkage to oil prices. Market price signals, which in a competitive Europe-wide market would reflect wholesale price volatility (or seasonality), are largely absent, so the incentive for private storage investors is much weaker than it could be. One could say the European liberalisation effort enhanced competition, but to a certain degree to the detriment of investment outcomes.

Given that the market is not yet functioning properly, regulation is of paramount importance. But supportive regulatory regimes for building storage in Europe are rare: although pipeline systems and storage facilities are inextricably associated, storage is not subject to regulated third-party access under the second EU gas directive. This directive regulates access to the distribution network, but access to storage facilities has to be negotiated individually. Coherent, competition-enhancing regulation is clearly called for. Such a regulation should create incentives for private investment in storage facilities, set rules for minimum storage capacity and aim at increasing both physical and economic interconnection of the European energy market. It should define rules for the case of emergency at a supranational level.[3] An expansion of the LNG infrastructure would also be crucial to increase both flexibility and diversity of gas sources.

J.A.

---

[1] For EU member countries it is "at least 90 days of average daily consumption in the preceding calendar year". (EU-Council Directive 2006/67 of 24 July 2006.) That strategic reserve is generally held either by industry or a combination of industry and a public entity and is supposed to help countries cope with severe supply disruptions. the individual national reserves can be bundled so that they can be made available to member states in case of a supply shock. This was the case when hurricane Katrina struck and devastated a large portion of oil production capacities in the Gulf of Mexico: the IEA member countries reacted within 10 hours.

[2] Across all IEA countries (Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Republic of Korea, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom, the United States) the variable cost of maintaining enough gas in strategic storage to satisfy a 90-day net import standard across the IEA is USD 5.4 billion per year.

[3] With the EU-Directive 2004/67/EC, which took effect in 2006, an important basis has been created. But this document does not contain rules for obligatory storage capacities, for example, it only states "the member states can implement minimum storage standards …".

EVALUATION OF PESTICIDE REGULATION

By passing a new pesticide regulation and extending the list of banned substances, the European Parliament has reignited the debate on effects of pesticide use on the environment and human health. Strong public concerns about this issue were first raised after the publication of Rachel Carson's "Silent Spring" in 1962. Thereafter, increased public consciousness led to the ban of the pesticide DDT – used mainly in agriculture – in the US in 1972. Since then, progress on banning toxic substances has been observed globally, but the efforts are not yet sufficient everywhere.

The use and export of pesticides were issues in multilateral environmental agreements (MEA), such as the Stockholm Convention on persistent organic pollutants and the Rotterdam Convention, both of which were ratified in 2004. MEAs are international treaties on transboundary environment problems that target on supranational solutions and oblige parties to implement respective laws. The Stockholm Convention bans the use of many toxic pollutants and specifies a list of twelve persistent organic pollutants (POPs) that move long distances in the environment. The Rotterdam Convention controls trade restrictions and regulations for toxic chemicals. It promotes proper labelling of exported hazardous chemicals and stipulates that exports of banned or restricted chemicals can only take place with the prior consent of the importing party.

Yale and Columbia University have constructed the Environmental Performance Index (EPI). The EPI quantifies the global progress towards higher ecosystem vitality, sound natural resource management and increased human health. The index combines several performance indicators. One is pesticide regulation, which is a particularly interesting variable because it evaluates policies and regulation in the field of environmental pollution.

The EPI is composed of two parts, and a total score of 22 can be reached per country. First, it tracks the political action taken with respect to the ratification of the two MEAs and their implementation into national law. A country receives one point for signing the convention and one point for submitting an implementation plan. Thus a total of 4 points can be gained from this section. Second, the index counts the number of banned pollutants within a country. The index focuses on a list of nine substances that are relevant to agriculture (aldrin, chlordane, DDT, dieldrin, endrin, heptachlor, hexachlorobenzene, mirex and toxaphene). Each one is assigned a maximum of two points. A banned pesticide receives a score of two, a restricted pesticide only a score of one. Thus, a country can obtain a score of 18 from banning pollutants.

The table shows a single country's absolute score as well as the proximity-to-target (PT) in percent. In the EPI for 2008 a total of 149 countries are ranked 21 countries reach the full score for pesticide regulation, twelve of which are European. Interestingly, amongst them the African countries Congo and Burundi are also ranked, although they belong to the poorest 20 percent of all ranked countries. About half of the countries have scores of at least 80 percent of

Table

Pesticide regulation

Rank	Country	Value	PT	Rank	Country	Value	PT
1	Armenia	22	100	26	Estonia	21	95.5
2	Australia	22	100	27	France	21	95.5
3	Austria	22	100	28	Greece	21	95.5
4	Bulgaria	22	100	29	Hungary	21	95.5
5	Burundi	22	100	30	Ireland	21	95.5
6	Canada	22	100	31	Italy	21	95.5
7	Chile	22	100	33	Latvia	21	95.5
8	Congo	22	100	34	Luxembourg	21	95.5
9	Czech Rep.	22	100	37	Netherlands	21	95.5
10	Denmark	22	100	41	Poland	21	95.5
11	Finland	22	100	42	Portugal	21	95.5
12	Germany	22	100	43	Spain	21	95.5
13	Japan	22	100	46	United Kingdom	21	95.5
14	Jordan	22	100	48	Brazil	20	90.9
15	Lithuania	22	100	49	Croatia	20	90.9
16	New Zealand	22	100	66	Slovenia	19	86.4
17	Norway	22	100	68	Turkey	19	86.4
18	Romania	22	100	69	United States	19	86.4
19	Slovakia	22	100	88	South Africa	14	63.6
20	Sweden	22	100	90	China	13	59.1
21	Switzerland	22	100	97	Macedonia	10	45.5
22	Belgium	21	95.5	111	India	3	13.6
24	Cyprus	21	95.5	145	Russian Fed.	0	0

Source: Esty, Daniel C., Marc Levy, Christine Kim, Alexander de Sherbinin, Tanja Srebotnjak, and Valentina Mara (2008), *2008 Environmental Performance Index*. New Haven, Yale Center for Environmental Law and Policy. Available at http://epi.yale.edu.

target-fulfilment. Among the countries with a score below 10 are important agricultural countries like Bangladesh, Pakistan, Russia and Taiwan, as well as a number of very poor countries. Some of the lowest scoring countries have signed both conventions, but have not yet eliminated any of the pollutants.

This index is particularly interesting for research questions dealing with the relationship between income and environmental protection or the impact of trade on the environment. The index consistently quantifies the regulatory progress and the awareness of governments for the transboundary problem of pollution.

J.K.

New at DICE Database

Recent entries to the DICE Database

In December 2008 and in the first quarter of 2009 the DICE Database received about 260 new entries, consisting partly of updates of existing entries and partly of new topics. Special point was the new establishment of the topic "Infrastructure" with several subfolders containing tables and charts on regulation policies, processes and policies in communication networks and transportation. Furthermore there are some new entries in the different folders as well as updates of existing entries. Some topics are mentioned below:

- Banking
- Budget Practices
- Health Care Systems
- Inflation Measures
- Integration of Immigrants
- Knowledge Creation and Innovation Performance
- Labour Migration
- Monetary Assistance to Families
- Pension System Characteristics
- Taxation of Labour.

Forthcoming Conferences

The Roland Coase Institute – Workshop on Institutional Analysis
10–15 May 2009, in Bratislava

CESifo Venice Summer Institute 2009 – Workshop on the Economics and Politics of Climate Change
6–7 July 2009, in Venice

The workshop "The Economics and Politics of Climate Change" takes stock of recent research on the economic effects of global warming, as well as analysing political support for various measures intended to reduce CO_2 and other emissions, or to absorb CO_2 through various measures, like reforestation.
Organisers: Panu Poutvaara, Mika Widgrén

Second Annual Research Symposium on The Economics and Law of the Entrepreneur at Northwestern University School of Law
11–12 June 2009, in Chicago
Organisers: Daniel F. Spulber, Henry N. Butler

Annual Conference of The Society for Economic Design (SED) 2009
22–24 June 2009, in Maastricht
Organisers: Hans Peters, Walter Trockel

EEA/ESEM, Joint Annual Meeting
23–27 August 2009, in Barcelona

Annual Conference of the European Association of Law and Economics
17–19 September 2009, in Rome

The focus of the conference will be on the economic analysis of ancient law, with a particular focus on Roman law.

New Books on Institutions

Restoring Financial Stability: How to Repair a Failed System
Viral V. Acharya and Matthew Richardson (eds)
New York University Stern School of Business
Wiley Finance, New York 2009

Growing Unequal? Income Distribution and Poverty in OECD Countries
OECD, Paris 2008

Corporate Governance Principles, Policies and Practices
Robert Tricker
Oxford University Press, Oxford 2009

Introduction to Modern Economic Growth
Daron Acemoglu
Princeton University Press, Princeton 2008

The Origin and Development of Financial Markets and Institutions. From the Seventeenth Century to the Present
Jeremy Atack and Larry Neal (eds)
Cambridge University Press, New York 2009

Online information services of the CESifo Group, Munich

Newsletter *ifo*

The Ifo Newsletter is a free service of the Ifo Institute and is sent by e-mail every month. It informs you (in German) about new research results, important publications, selected events, personal news, upcoming dates and many more items from the Ifo Institute.
If you wish to subscribe to the Ifo Newsletter, please e-mail us at: newsletter@ifo.de.

CESifo THE CESIFO NEWSLETTER

CESifo publishes about 20 working papers monthly with research results of its worldwide academic network. The CESifo Newsletter presents selected working papers (in English) in an easily understandable style with the goal of making its research output accessible to a broader public.
If you wish to subscribe to the CESifo Newsletter, please e-mail us at: saavedra@cesifo.de.

If you wish to receive our current press releases, please e-mail us at: presseabteilung@ifo.de.

You can also request these services by fax:

Ifo Institute for Economic Research, fax: (089) 9224-1267

Please include me in your mailing list for:
☐ Ifo Newsletter ☐ CESifo Newsletter ☐ Ifo Press Releases

Name: ....
Institution: ....
Street: ....
City: ....
Telephone: ....
Fax: ....
E-mail: ....

DICE
Database for Institutional Comparisons in Europe
www.cesifo-group.de/DICE

The database DICE was created to stimulate the political and academic discussion on institutional and economic policy reforms. For this purpose, DICE provides country-comparative information on institutions, regulations and the conduct of economic policy.

To date, the following main topics are covered: Business and Financial Markets, Education and Innovation, Energy and Natural Environment, Labour Market and Migration, Public Sector, Social Policy, Values. Information about Basic Country Characteristics is provided for the convenience of the user.

The information of the database comes mainly in the form of tables – with countries as the first column – but DICE contains also several graphs and short reports. In most tables, all 27 EU and some important non-EU countries are covered.

DICE consists primarily of information which is – in principle – also available elsewhere but often not easily attainable. We provide a very convenient access for the user, the presentation is systematic and the main focus is truly on institutions, regulations and economic policy conduct. Some tables are based on empirical institutional research by Ifo and CESifo colleagues as well as the DICE staff.

DICE is a free access database.

Critical remarks and recommendations are always welcome.
Please address them to
ochel@ifo.de
or
hoffmann@ifo.de
or
rohwer@ifo.de