April 11, 2005
(revised September 15, 2005)

Labour Market 'Reform' and Cross-Country Employment Performance: Does the Evidence Tell an Orthodox Tale?

Dean Baker, Andrew Glyn, David R. Howell,
and John Schmitt

LoWER Annual Conference 2005
Mannheim, April 15-16

It has become widely accepted that the cross-country pattern of employment performance, measured usually by the unemployment rate, is explained largely, or even entirely, by the relative rigidity of national labor markets.[1] There are two leading rigidity stories: that high *levels* of social protection have limited the ability of economies to adjust to shocks (Blanchard and Wolfers, 2000), and that *changes* in protective labor market institutions explain employment performance over time and across countries, since these changes either promote or inhibit necessary wage and employment adjustments (e.g., Nickell et al., 2002; IMF, 2003; Nickell, 2003). Rigidity explanations have long been a central tenet of orthodox economics, but there is an equally long, if less influential, dissenting position, most famously illustrated by Keynes' attack on the wage rigidity explanation for unemployment in the Great Depression.

What has distinguished the last decade on this question is the growing influence of cross-country statistical studies that employ measures of various labor market institutions. We have addressed the robustness of this claim in previous papers. In Baker et. al (2005), our survey of the most influential cross-country studies to date found that, while they generally conclude that the statistical evidence offers reasonably strong support for the orthodox view, they are far from unanimous in their estimates of the economic and statistical significance, and in some cases even the direction, of the effects of standard institutional variables on unemployment. Indeed, a number of the prominent papers explicitly refer to this lack of robustness of their own results across specifications and variable definitions. In our own tests using 5 year periods and modeled after those published by Nickell (1997), we found that the strong cross-sectional relation between unemployment and institutions found by Nickell (1997) for the mid-1980s to the mid-

1990s disappeared entirely with the substitution of improved OECD measures of the same labor market institutions.

In another paper (Baker et al., 2005), we focused on the robustness of the conclusions drawn by a recent study by the IMF. Following Nickell et al. (2003), this IMF report relies on annual data, which we argue is problematic on several counts – most importantly the dubious quality of the institutional variables, which are difficult enough to measure over 5 year periods (the best sources are often periodic estimates by the OECD, which must then be annualized by interpolation). After replicating the IMF results, we found them highly sensitive to minor changes in specification – changes that are quite consistent with standard econometric approaches in this literature. Indeed, contrary to the impression given in the IMF text, there are important differences in estimates of effects even across their own published regressions.

In addition to this econometric evidence, in recent years the orthodox case has been made with simple, essentially bivariate, associations between summary indicators of labor market reforms and changes in unemployment for the 1990s. In this paper, after presenting some institution-unemployment scatter plots of own, we focus the quality of the evidence in two prominent examples of the use of this sort of evidence.

It is commonly argued that the strictness of employment protection and the generosity of unemployment benefits can, separately, explain much of the difference between the best and worst performing labor markets. The first section takes a look at the evidence for simple correlations between unemployment levels and measures of these institutions since the early 1980s. The scatter plots for 20 countries over 4 five-year periods should show high unemployment for those countries with high levels as well as upward changes in these measures (and low unemployment for low levels and downward changes).

Sections 2 and 3 assess two influential attempts to use composite indicators of labor market 'reforms' to account for the cross-country pattern of unemployment. Section 2 examines Nickell's (2003) "ticks and crosses" analysis. While a rather casual exercise, it has been quite influential, both for its prominent author and very strong conclusions;it was the main reference to empirical evidence in a recent *Journal of Economic Perspectives* article that pronounced that the pattern of unemployment across Europe could be explained by the pattern of labor market rigidities: "Nickell (2003) summarizes

these diverging experiences by correlating the change in unemployment across countries in the 1990s with labor market reforms and finds the expected sign. Therefore, evidence supports the traditional view that rigidities that reduce competition in labor markets are typically responsible for high unemployment. Reducing these rigidities across the board seems to work" (St. Paul, 2004p. 53).

Section 3 considers the evidence presented by to the OECD's "Implementing the Jobs Study" report, a follow-up to the OECD's extremely influential *Jobs Study*, which also finds a strong association between a composite measure of labor market reforms and unemployment across OECD countries. We conclude in section 4.

1. Labour Market Institutions and Cross-Country Employment Performance: the Simple Statistical Evidence

It is frequently argued that employment performance will improve with the reform of each "employment-unfriendly" labor market institution. Indeed, this premise lies behind both the OECD's (1999) reforms/NAIRU scatter plot and Nickell's (2003) ticks/crosses analysis, both of which are considered in detail below. In Baker et al. (2005), we produced six scatterplots of labor market institutions against unemployment for 20 countries and 4 periods (1980-84, 1985-99, 1990-94, 1995-99). We found little, if any, correlation between these standard institutional measures and unemployment. We reproduce three here.

Figure 1 shows the plot of country-time points for the strictness of employment protection laws and unemployment. An OECD survey on employment protection (1999; see also OECD, 2004) found that "empirical results are somewhat mixed… Bertola (1992), Nickell and Layard (1997), and OECD (1999b) were unable to find a statistically significant relationship between EPL and the unemployment rate…". This is, indeed, precisely what Figure 3 indicates.

Despite both the widely accepted view that unemployment benefit generosity lies at the heart of the unemployment problem and the likelihood of at least some reverse causation, Figures 2 and 3 show virtually no association between the standard measures of unemployment benefit generosity and unemployment over the 1980-99 period. Figure 2 shows a slight positive (but statistically insignificant) relationship between the unemployment rate and the replacement rate due to a single outlier country, Spain.

Directly below the four Spanish points are those for Sweden; while both countries had similar replacement rates (ranging from 65-76%), the five-year average unemployment rates in Spain ranged from 16-20 percent while Swedish unemployment rates ranged from 2-8 percent. Another example of the lack of correspondence between replacement rates and unemployment can be seen with France and the Netherlands. While French replacement rates were about 58 percent from 1980-99 and Dutch rates were much higher(70 percent), French unemployment *rose* from 8 to 12 percent while Dutch unemployment *fell* from 8 to 5 percent.

Figure 3 shows that there is also no simple association between unemployment benefit duration and unemployment levels across these 20 countries and 4 time periods. With similar unemployment rates, New Zealand (1.04) and the U.S. (.15 -.19) are at opposite ends of the spectrum on this measure of duration. On the other hand, the quintessential welfare state, Sweden, with a strong commitment to active labor market policies (training and job placement services), gets a duration score (.04-.05) that is far smaller than that of even the United States. Spain's duration score since 1985 (.25-.28) is only slightly above that of the United States, but far below that of the United Kingdom (.70-.73); nevertheless, Spain has had unemployment rates 2-3 times higher than the United Kingdom (20.1 vs 7.3 for 1995-99).

2. Ticks, Crosses, and Unemployment

Stephen Nickell has helped pioneer the panel data analysis labor market institutions and employment performance across developed countries in a literature that used methods that became increasingly sophisticated and institutional measures that improved in quality over the course of the 1990s. In "Labour Market Institutions and Unemployment in OECD Countries," he turns back to a much simpler, but perhaps even more powerful analysis that develops a labor market reforms scorecard and relates it to the change in unemployment between the 1980s and the end of the 1990s.

Nickell makes two empirical claims. The first is that the problem of European unemployment today is concentrated in "the big four": France, Germany, Italy and Spain. This is a very simple but important point. A good explanation of the "European" unemployment problem must explain the persistence of high unemployment in these four big European countries. As he pointed out in his 1997 *Journal of Economic Perspectives*

paper, much of Europe lives in regions with unemployment rates that are lower than the U.S. rate.

But the analysis in the paper is dedicated to the much bolder second claim, which is that changes in unemployment can be well-explained by changes in (mostly "bad") labor market institutions. Nickell assesses the impact of deregulation, or "labor market reforms", by assigning "ticks" to "good" changes (employment friendly reforms) and "crosses" to "bad" ones, for 9 measures and 20 countries from the early 1980s to the late 1990s. Nickell's scorecard of labor market policy reforms accounts for just over half the variation in unemployment trends across 19 OECD countries since the early 1980s, nearly identical to what his most recent and most sophisticated regression analysis produced (Nickell et al., 2002/3). And as noted above, this reforms scorecard evidence has already been extremely influential (St. Paul, 2004).

But how reliable is this scorecard evidence? Leaving aside the difficulty of determining what threshold should determine a cross or a tick for each of the 9 measures, we begin with some questions raised by the interdependence of these institutions and our ability to measure them.

1. With three of the nine measures, unemployment benefit generosity is heavily weighted. But given the interdependence of these dimensions of the system – the first year replacement rate, duration of benefits, and eligibility strictness – allocating ticks and crosses to any one in isolation can be misleading. Thus, the replacement rate might be relatively high, but offset with very low duration (Sweden). Making matters even more problematic, the way the duration measure is defined ensures that it will vary with changes in the replacement measure: falling replacement rates (the rate for year 1) mean that, all else equal, duration generosity automatically increases (since the latter is measured as the replacement rates in years 2-5 relative to the first year). Ireland is a good example.

2. Another dimension of benefit generosity is the strictness of benefit eligibility. There is no good cross-country measure of eligibility, much less its change over time. Nickell uses a measure from a Danish study for the mid-1990s and allocates ticks on the basis of "author's judgment."

Furthermore, Nickell's conclusion that over half of the variation in unemployment is explained by his ticks and crosses is quite fragile. Even given the distribution of ticks and crosses, every alternative base and end year and mix of countries we tried produced weaker results. The second row of Table 1 shows that the explanatory

power of the equation collapses without the Netherlands, the UK, and Ireland. That is, for 17 or the 20 countries, neither ticks nor crosses are significantly related to the change in unemployment. Row 3 shows that if the base year is changed to 1980-81 from 1980-87, only the ticks remain a statistically significant factor, but the explanatory power remains fairly high (.44). We also explored the effects of Nickell's reform scorecard on the employment rate. After all, the logic of the labor market reform prescription is that greater flexibility and lower levels of benefits will spur employer demand and provide greater incentives for workers to work, so the balance of ticks and crosses should also do a good job of explaining changes in the employment rate. Row 5 shows that this expectation is weakly supported for the two decade period (ticks are barely significant with an R2 of .25). But row 6 shows that there is no meaningful relationship without the Netherlands: for 19 of these 20 countries, Nickell's policy reform scorecard provides no explanatory power for changes in the employment rate.

Third, the ticks/crosses scorecard does not identify the high unemployment countries. Nickell's first empirical proposition – that unemployment in Europe is essentially a problem of 4 large countries – would be consistent with his second: that his policy reforms scorecard will pick out the poor performers. Indeed, based on the ticks/crosses analysis, Nickell states that "We may reasonably conclude that the countries which had very high unemployment in the early 1980s and still have high unemployment today simply have too few ticks and/or too many crosses." So the big four countries with high persistent unemployment should be located at the negative end of the spectrum (many more crosses than ticks). Figure 4 reports the net ticks/crosses. In fact, only France fits this prediction.[2] But with a reforms scorecard balance of -3, France (persistent high unemployment) exactly matches Switzerland (persistent low unemployment). High unemployment Germany and Spain get the same score (+1), which is far superior to the score given Austria (-2), a persistent low unemployment country. This figure makes clear that Nickell's is not a reasonable conclusion - ticks and crosses do not, in fact, identify countries with persistent high unemployment since the early 1980s.

In sum, this is not very compelling evidence for the orthodox labor market reforms prescription. To be fair, Nickell recognizes that this is rather soft, qualitative evidence, and suggests that "Readers who prefer panel data analysis can consult the papers discussed in the second section." We would suggest that readers who would like a

detailed critical assessment of this panel data analysis literature, one that underscores how fragile the results are, might also consult Baker et al (2005).

3. The OECD's NAIRU/Reforms Evidence

A central pillar of OECD labor market policy has been that reforms that reduce labor market rigidities are the answer to persistent high unemployment. An enumeration of such reforms was carried out by the OECD as part of its follow-up to *The Jobs Study* (OECD 1994). Their 1999 survey (OECD 1999b) provides an extremely comprehensive listing of changes in the generosity of unemployment benefits, the strictness of employment protection laws, the level of minimum wages and the like, focused on the period from 1995 but also with summary data from the early 1990s. The OECD listed all the reforms suggested for each country in its labor market reviews, developed a weighting system for assessing their significance, and then analyzed whether the recommended reform had been fully implemented, partially implemented, ignored, or even flouted (in the sense that policy had moved in the "wrong" direction).

The OECD found a significant positive relation between their measure of "follow-through" by countries in response to OECD recommendations and the extent to which the NAIRU fell in the 1990s (OECD 1999b: figure 2.7). But such a measure ignores the very different number of recommendations for labor market reforms that countries received from the OECD (varying from 4 in the case of USA and Australia to 21 for Finland and 23 for Germany). The effect of reforms on unemployment should presumably depend on how many were implemented, not simply the *proportion* of OECD suggestions which were followed. Accordingly we constructed an alternative index showing the "volume" of labor market deregulation recommendations that were actually carried out, which depends on both the number of measures advocated by the OECD and their "follow-through" by the countries (see Data Appendix). We focused our index on reforms connected with the benefits, employment protection, and wage bargaining systems, as these constitute the core of labor market deregulation.

Figure 5 compares this index of labor market deregulation in the 1990s with the OECD's estimate of the change in the NAIRU over the same period for 21 OECD member countries. It is clear that there is no significant relationship between these measures of deregulation and the change in unemployment across OECD countries.

Obviously Ireland is an extreme case with the most dramatic fall in the NAIRU accompanied by rather little labor market reform. However even if Ireland is excluded (and this would be hard to justify), the relationship between deregulation efforts and structural unemployment across countries still appears very weak (only about one tenth of the variance in the NAIRU change being "explained"). By this measure, changes in structural unemployment which have occurred across the major OECD member countries in the 1990s are not systematically associated with the extent of labor market reform.

4. Conclusion

This paper has argued against the view that changes in structural unemployment in OECD countries over the past two decades has been dominated by the extent to which countries have implemented labour market reforms. There are many problems of measurement involved in evaluating the effect of changes in labour market institutions as our discussion of two studies focused on this issue confirms. Thus far, robust evidence of systematic effects of labour market reforms has not been produced, though this has not dimmed the confidence with which such reforms are promoted.

References

Baker, Dean, Andrew Glyn, David R. Howell, and John Schmitt (2003), "Unemployment and Labor Market Institutions: The Failure of the Empirical Case for Deregulation," Report to the International Labour Organization.

Baker, Dean, Andrew Glyn, David R. Howell, and John Schmitt (2005) "Labor Market Institutions and Unemployment: A Critical Assessment of the Cross-Country Evidence", in David R. Howell, ed., *Fighting Unemployment: The Limits of Free Market Orthodoxy*, (Oxford University Press).

Blanchard, Olivier and Justin Wolfers (2000). "The Role of Shocks and Institutions in the Rise of European Unemployment: the Aggregate Evidence." *The Economic Journal* 110 (March): C1-C33.

International Monetary Fund IMF (2003), "Unemployment and Labor Market Institutions: Why Reforms Pay Off,' *World Economic Outlook* (April) pp. 129-150.

Nickell, S, L.Nunziata, W.Ochel, G.Quitini (2002), "*The Beveridge Curve, Unemployment and Wages in the OECD from the 1960s to the 1990s," (CEP, LSE: London various versions*) and published in P.Aghion et al (eds.) *Knowledge, Information and Expectations in Modern Macroeconomics* (Princeton University Press), 2003.

Nickell, Stephen (2003), "Labour Market Institutions and Unemployment in OECD Countries," CESifo DICE Report 2/2003.

St. Paul, Gilles (2004), "Why are European Countries Diverging in their Unemployment Experience?," *Journal of Economic Perspectives,* vol. 18, no. 4 (fall), pp. 49-68.

Appendix: Measurement of Labor Market Deregulation

The OECD's report, *Implementing the Jobs Study* (1999b), lists five areas of labor market reform: including unemployment and non-employment benefits (12 sub-sections including replacement rates, duration, and eligibility), wage formation (6 sub-sections including bargaining decentralization, minimum wages) and EPL and working time arrangements (10 sub-sections including authorization for dismissals, constraints on part-time work). They also include active labor market policies (ALMP) and the tax wedge. We did not include the latter two areas in our analysis. ALMP does not really fall in to the heading of labor market deregulation (its "active" nature reflects the need to overcome market failures rather than widening the scope for market forces). The overall tax wedge, while it may be important for employment, can hardly be considered to be determined primarily on labor market grounds and it seems arbitrary to examine it in relation to the rather few countries on the receiving end of OECD recommendations for tax wedge reduction. We did, however, include in our analysis two minor elements classified under the tax wedge - targeted reductions in social insurance and the taxation of low income earners - as these are explicitly aimed at increasing wage flexibility by "making work pay". The importance of each of the 30 policy sub-categories was weighted following the OECD's rating of their importance. We used a numerical version of this weighting constructed by Van Ploek and Borghijs (2001) with minor modifications. The country's response to the OECD's suggestions (weighted 1.0 for "sufficient action", 0.5 for "more action needed", 0 for "no action" and -0.5 for "opposite action") were taken from OECD's Appendix tables (1999b). A country which had been recommended to implement every one of the 30 policies and which had fully carried them out would have an index of 0.64. Thus a low number for the index reflects either few recommendations or a low rate of compliance.

Table 1: Effects of Nickell's "Good" and "Bad" Labour Market Policy Changes
Using Alternative Measures of Employment Performance and Countries
(t statistics in parentheses; t >2.0 are highlighted)

	Ticks	Crosses	R2	N
1. Chg U Rate: 1980-87: 2000-01 (Nickell)	**-1.25 (3.1)**	**1.21 (2.2)**	.51	20
2. Chg U Rate: 1980-87:2000-01 (w/o NL, IRE, UK)	-.74 (1.4)	.9 (1.62)	.24	17
4. Chg U Rate: 1980-81: 2000-01	**-.96 (2.9)**	.73 (1.6)	.44	20
5. Chg EPOP Rate: 1980-81: 2000-01	**1.68 (2.1)**	-.86 (.76)	.25	20
6. Chg Epop Rate: 1980-81: 2000-01 (w/o NL)	1.06 (1.4)	-.02 (.0)	.10	19

Data: Unempl change from 1980-87 to 2000-01 and the numbers of "ticks" and "crosses" are taken from Nickell (2003). The other unemployment and the employment rates are from the OECD.

Figure 1: Employment Protection Laws and Unemployment, 1980-99
(20 countries, 4 five-year periods)

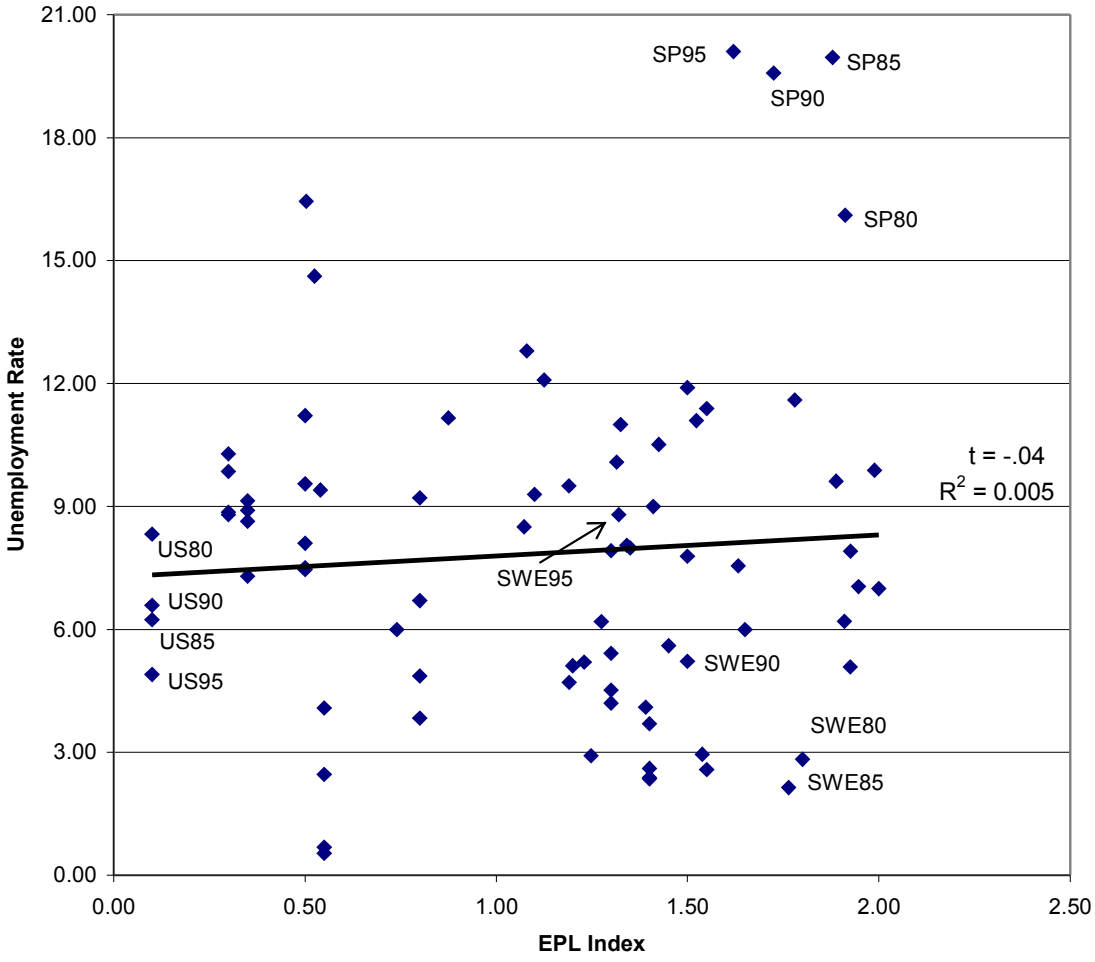


Figure 2: The Unemployment Benefit Replacement Rate and Unemployment, 1980-99
(20 countries, 4 five-year periods)

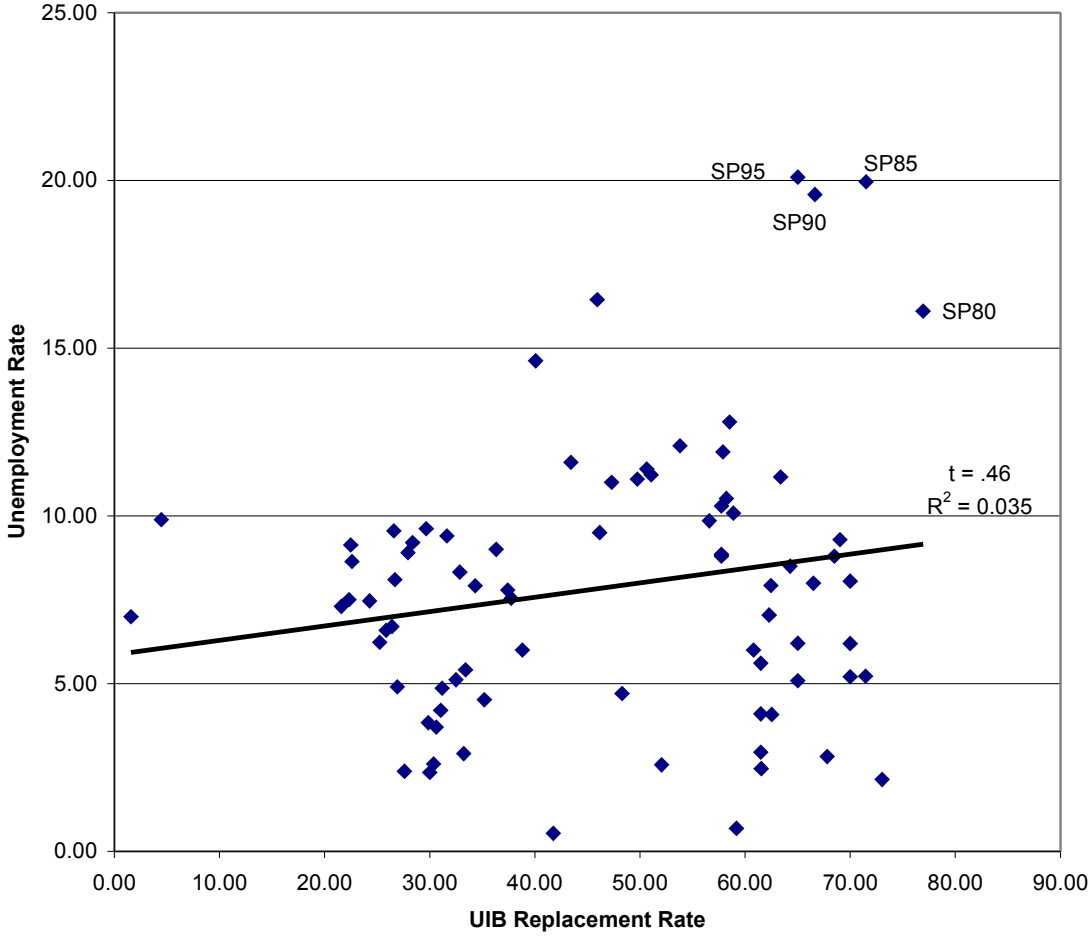


Fig 3: Unemployment Benefit Duration and Unemployment, 1980-99
(20 countries, 4 five-year periods)

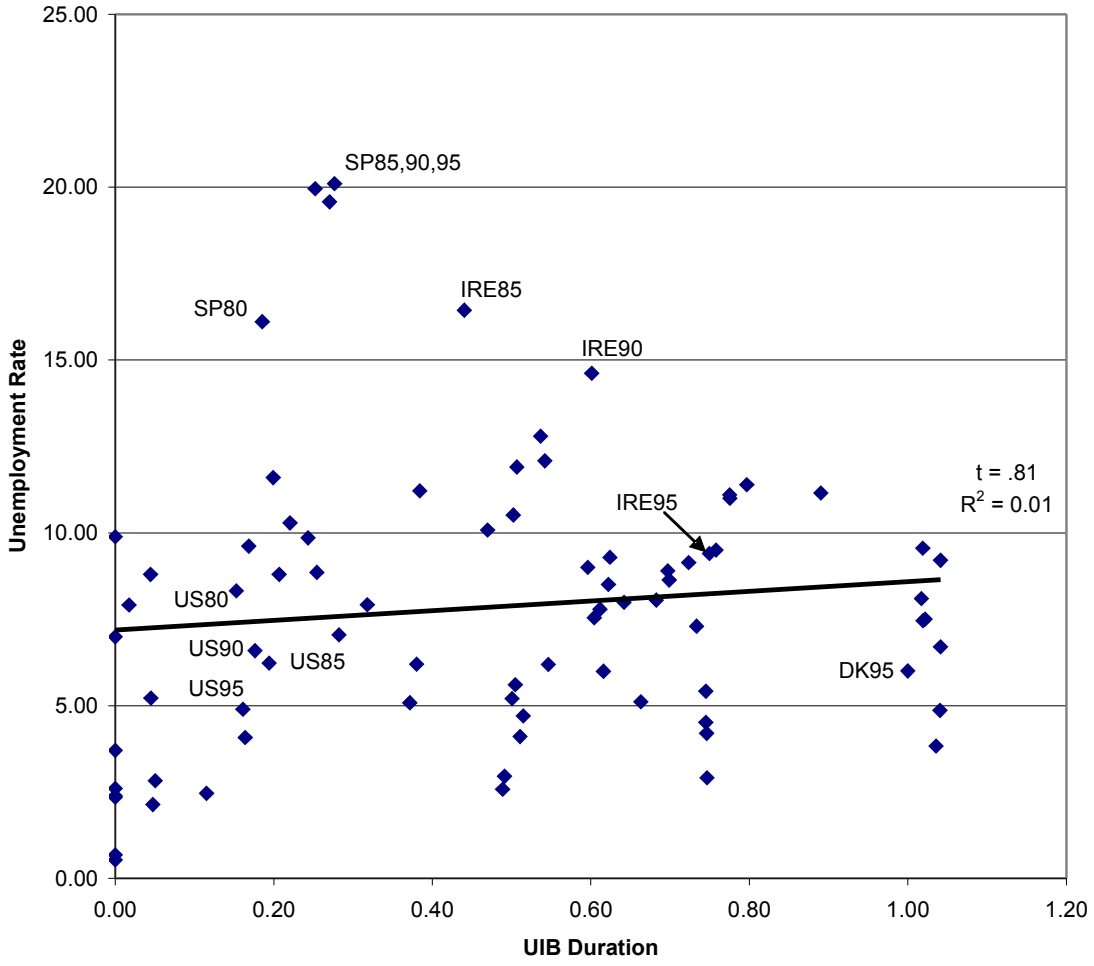


Figure 4: Nickell's Policy Reform Scorecard:
Net "Good" and "Bad" Labour Market Policy Changes,
Early 1980s to Late 1990s

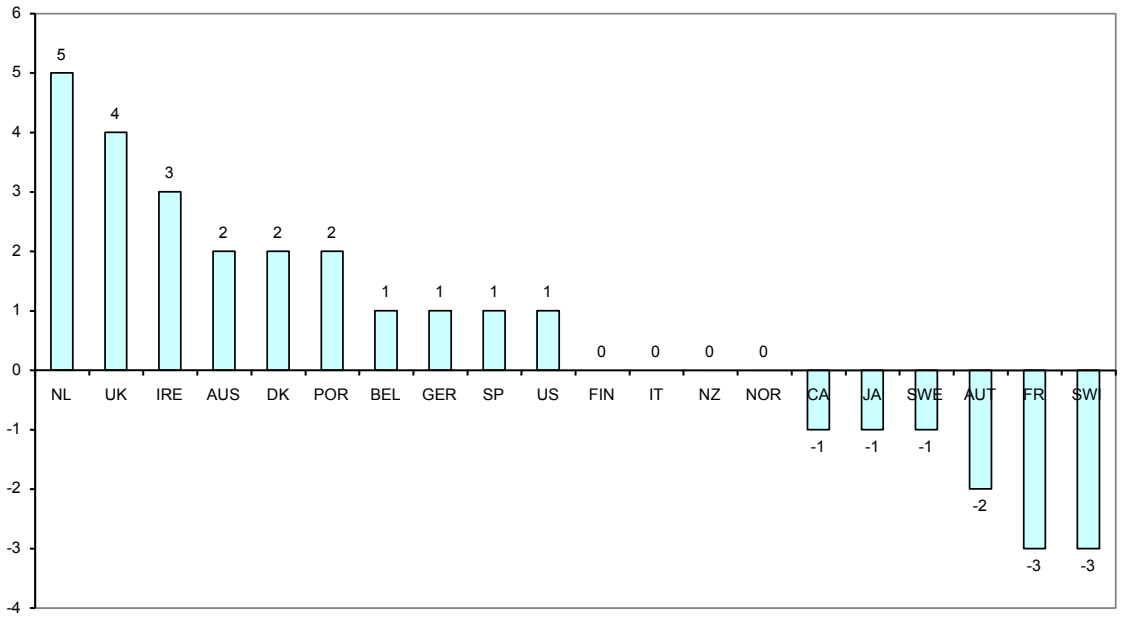
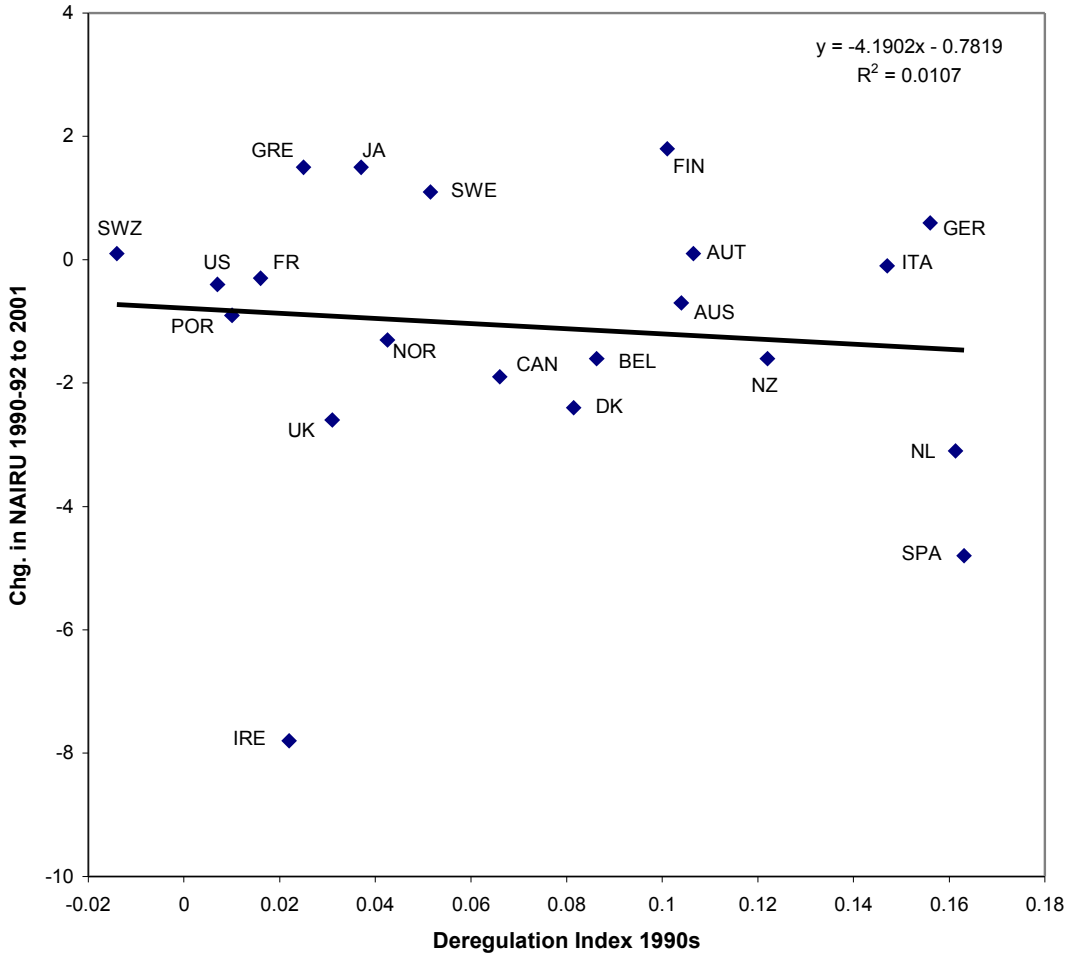


Figure 5: Labor Market Deregulation and Changes in the NAIRU for 21 OECD Countries in the 1990s



[1] For example, Gilles St. Paul (2004: 51) writes that in the later 1980s and 1990s "a rough consensus emerged that high unemployment in Europe was due to labor market rigidities."

[2] It could easily be argued that Nickell's allocation of ticks and crosses for France is among the most problematic. For example, based on Nickell's table 5, France gets a cross based on a modest increase in the 1980s, while the 1990s actually show a slight decline. Again, on union coverage, France gets a cross on the basis of changes that took place in the 1980s, not the 1990s, and gets no credit (a tick) for reducing union density from 16% to 10%, a level below that of the U.S.! It gets another for increasing strictness of employment protection, which was entirely due to the increase for the 15% of the workforce who are temporary workers. Interestingly, Nickell's criterion (a rise of more than .1) gives France a cross, even though it takes a decline of .2 to get a tick, and even though France's score changes from 1.3 to 1.4 (exactly .1). Both the rules of the game and the allocation seem not in France's favor!