Revisiting Hysteresis in Unemployment for Ten European Countries: Panel Unit Root Tests

Tsangyao Chang*
Department and Graduate Institute of Economics
Feng Chia University, Taichung, Taiwan

Hann Wen Tzeng
Department of Banking & Risk Management
The Overseas Chinese Institute of Technology
Taichung, Taiwan

ABSTRACT

The hysteresis hypothesis in unemployment for ten European countries is tested using the Levin and Lin (1992), Im et al., (1997), Taylor and Sarno (1998) and Breuer et al., (2001) panel-based unit root tests for the 1961-1999 period. The results from the first three tests provide strong evidence in support of the hysteresis hypothesis given the European countries' unemployment data. The hypothesis is also confirmed for all the European countries except Belgium and the Netherlands when Breuer et al.'s (2001) SURADF test is conducted.

Keywords: Hysteresis in Unemployment; European Countries; Panel Unit Root Tests

JEL classification: C22; C23

---

*Corresponding author: Professor, Department and Graduate Institute of Economics, Feng Chia University, Taichung, Taiwan. TEL: 886- 4-2451-7250 ext. 4484. FAX: 886-4-2451-8737. E-mail: tychang@fcu.edu.tw.

1. INTRODUCTION

Confirming the validity of the hypothesis of hysteresis in unemployment is critical for both empirical researchers and policymakers. Considering the assumptions inherent in the hysteresis hypothesis in unemployment, if unemployment is the I(1) process, then the shocks affecting the series will have permanent effects, thus shifting the unemployment equilibrium from one level to another. Should this be the case, from the policy perspective, policy action is, indeed, required to return unemployment to its original level. On the other hand, if unemployment is the I(0) process, the effects of the shock will merely be transitory, making the need for policy action less mandatory since unemployment will eventually return to its equilibrium level. The I(0) process has commonly been referred to as the natural-rate of unemployment hypothesis (NAIRU) for it characterizes unemployment dynamics as a mean reversion process.

Because hysteresis is associated with non-stationary unemployment rates, unit root tests have widely been used to investigate its validity. Using 1853-1984 data for France, Germany, the United Kingdom and United States, Blanchard and Summers (1986) laid the groundwork by employing conventional unit root tests to investigate the effects of hysteresis on unemployment, and they were unable to reject the non-stationarity of unemployment rates, except for the Untied States where they did find evidence of stationarity. A little later, Brunello (1990), using 1955-1987 Japanese unemployment data, was also unable to reject the null hypothesis of a unit root. Mitchell (1993) later adopted Perron's (1989) unit root test, which assumes one exogenously given structural break, and this similarly provided support for hysteresis in several OECD countries. Likewise, Jaeger and Parkinson (1994) reported that unemployment hysteresis exists in Germany, the United Kingdom and Canada, but not in the United States. Using 1970-1994 data, Roed (1996) reported on unemployment hysteresis in 16 OECD countries and the strong likelihood of it in Australia, Canada, Japan and several European countries, but like other researchers rejected it in the United States.

While these findings generally supported a unit root in unemployment and, therefore hysteresis, critics have claimed that the drawing of such conclusions may be attributed to the lower power of the conventional unit root tests employed. More recently, in fact, it has been reported that conventional unit root tests not only fail to consider information across regions, thereby leading to less efficient estimations, but also have low power against near-unit-root but stationary alternatives. It is not surprising that these factors should expectedly have cast considerable doubt on many of the earlier findings of a unit root in unemployment rates.

One proposed approach to increasing power in testing for a unit root involves the use of panel data. Levin and Lin (1992) and Im et al., (1997) developed the asymptotic theory and the finite-sample properties of ADF tests of panel data, and both have demonstrated that even relatively small panels yield large improvements with respect to power. These panel-based unit root tests are now being extensively used in empirical testing – particularly as found in the literature for purchasing power parity; for example, see MacDonald (1996), Oh (1996), Wu (1996), Papell (1997), Papell and Theodoridis (2001), and Wu and Wu (2001). As for unemployment, on testing the hysteresis hypothesis in unemployment for 48 contiguous U.S. states and 16 OECD countries by simultaneously using the univariate and the panel-based unit root tests of Levin and Lin (1992), respectively, Song and Wu (1997, 1998) observed that with the application of the standard ADF and P-P tests to individual unemployment series, the unit root null is never rejected. By sharp contrast, with data pooled for the panel-based unit root test, the unit root null can generally be rejected. Simply put, they found no support whatsoever for the hysteresis hypothesis. However, from their application of Im et al.'s (1997) panel-based unit root test for hysteresis in unemployment, what Leon-Ledesam (2002) concluded is that hysteresis for the EU and the natural-rate for the US are the most plausible hypotheses.

This study contributes to this line of research by determining whether hysteresis in unemployment is characteristic of the European labor market. We test the hysteresis hypothesis in unemployment for 10 European country data sets using the Levin and Lin (1992), Im et al., (1997), Taylor and Sarno (1998) and Breuer et al., (2001) panel-based unit root tests.

The remainder of this study is organized as follows. Section 2 presents the data used, and Section 3 describes the methodology used. Section 4 discusses the empirical findings and policy implications. Finally, Section 5 presents some concluding remarks.

2. DATA

This study employs the 1961-1999 unemployment rates for ten European countries, namely Belgium, Denmark, France, Ireland, Italy, the Netherlands, Portugal, the UK, Norway and Finland. All the data are from the AREMOS database of the Ministry of Education of Taiwan, and summary statistics are given in Table 1. The unemployment data indicate that Ireland and Norway have the highest and lowest average unemployment rates, respectively. The Jarque-Bera test results meanwhile indicate that, except for Norway and Finland, all the unemployment data sets are approximately normal.

3. PANEL UNIT ROOT METHODOLOGY

3.1. Levin and Lin (1992), Im, Pesaran and Shin (1997) and Taylor and Sarno's (1998) Panel-Based Tests

The conventional ADF test for single-equation unemployment is based on the following regression equation:

$$\Delta X_{it} = \alpha_i + \beta_i X_{i,t-1} + \gamma_i t + \sum_{j=1}^{k} \theta_{ij} \Delta X_{i,t-j} + \varepsilon_{it}, \quad [1]$$

where Δ is the first difference operator, X_{it} is the unemployment rate, ε_{it} is a white-noise disturbance term with a variance of σ^2, and t = 1, 2,…., T indexes time. The unit root null hypothesis of $\beta_i = 0$ is tested against the one-side alternative hypothesis of $\beta_i < 0$, which corresponds to X_{it} being stationary. The test is based on the test statistic $t_{\beta_i} = \hat{\beta}_i / se(\hat{\beta}_i)$ (where $\hat{\beta}_i$ is the OLS estimate of β_i in Equation [1] and $se(\hat{\beta}_i)$ is its standard error) since the single-equation ADF test may have low power when the data are generated by a near-unit-root but stationary process. Levin and Lin (1992, hereafter, Levin-Lin), on finding that the panel approach substantially increases power in finite samples when compared with the single-equation ADF test, proposed a panel-based version of Equation [1] that restricts $\hat{\beta}_i$ by keeping it identical across cross-sectional regions as follows:

$$\Delta X_{it} = \alpha_i + \beta X_{i,t-1} + \gamma_i t + \sum_{j=1}^{k} \theta_{ij} \Delta X_{i,t-j} + \varepsilon_{it}, \qquad [2]$$

where i =1,2,…N indexes across cross-sectional regions. Levin-Lin tested the null hypothesis of $\beta_1 = \beta_2 = ... = \beta = 0$ against the alternative of $\beta_1 = \beta_2 = ... = \beta < 0$, with the test based on the test statistic $t_{\beta} = \hat{\beta} / se(\hat{\beta})$ (where $\hat{\beta}$ is the OLS estimate of β in Equation [2], and $se(\hat{\beta})$ is its standard error).

While the Levin-Lin panel-based unit root test has become increasingly popular in applied work, one drawback is that β is restricted by being kept identical across regions under both the null and alternative hypotheses. Im et al., (1997, hereafter Im-Pesaran-Shin) relaxed the assumption of the identical first-order autoregressive coefficients of the Levin-Lin test and developed a panel-based unit root test that allows β to vary across regions under the alternative hypothesis. Im-Pesaran-Shin tested the null hypothesis of $\beta_1 = \beta_2 = ... = 0$ against the alternative of $\beta_i < 0,$ for some i.

The Im-Pesaran-Shin test is based on the mean group approach. They use the average of the t_{β_i} statistics from Equation [1] to perform the following t-bar statistic:

$$\overline{Z} = \sqrt{N}[\bar{t} - E(\bar{t})]/\sqrt{Var(\bar{t})} \qquad [3]$$

where $\bar{t} = (1/N)\sum_{i=1}^{N} t_{\beta_i}$, $E(\bar{t})$ and $Var(\bar{t})$ are respectively the mean and variance of each t_{β_i} statistic, and they are generated by simulations (for further details, see Im et al., 1997). This $\overline{Z}$ converges to a standard normal distribution. Based on Monte Carlo experiment results, Im et al., (1997) demonstrated their test is often even more powerful than that of the Levin-Lin panel test in finite samples. Even so, the problem of cross-sectional dependence is inherent in both the Levin-Lin and Im-Pesaran-Shin panel-based unit root tests. O'Connell (1998) has in fact shown that the true size of both tests can be far greater than the normal size when the underlying data-generating process (DGP) is characterized by $\text{cov}(\varepsilon_{it}, \varepsilon_{jt}) \neq 0$ for $i \neq j$. Though Levin and Lin (1992) and Im et al., (1997) both proposed controlling for cross-sectional dependence by subtracting the cross-sectional means before performing estimations of Equation [1] in order to remove the effect of a common time component, Mark (2001) has indicated that if common time effects are generated by a multi-factor process, then transforming the observations by subtracting the cross-sectional means will still leave some residual dependence across individuals. Such residual cross-sectional dependence has the potential to generate errors and lead to faulty inferences. O'Connell (1998) has also shown that the same procedure does little to reduce cross-sectional dependence and size distortions when the time component varies across regions. A straightforward way to handle cross-sectional dependence that may vary across regions is to estimate Equation [1] using the seemingly unrelated regression (SUR) estimator (Zellner, 1962). O'Connell (1998) found that size distortions can be avoided with little loss of power by basing the panel-based test on the SUR rather than the OLS estimation of Equation

[1]. However, the SUR panel-based test like the Levin-Lin panel-based test also has the drawback of restricting β by keeping it identical across regions under the alternative hypothesis. In light of this, Taylor and Sarno (1998) suggested a modified version of the SUR panel-based test, one that allows for different β values under the alternative hypothesis and controls for cross-sectional dependence. Taylor and Sarno (1998) called this the MADF test, and it is based on the SUR estimation of Equation [1] for $i = 1,2,...,N$. They also noted that this test is quite powerful in finite samples for the Monte Carlo experiments that they performed. To check the robustness of our results, we conduct all three of these panel-based tests in our study.

3.2. Breuer, McNown and Wallace's (2001) Seemingly Unrelated Regressions Augmented Dickey-Fuller Test (SURADF)

Breuer et al., (2001) showed the recent methodological refinements of the Levin and Lin test fail to fully address the "all-or-nothing" nature of the test. It is true that Im et al. (1997), Maddala and Wu (1997) and Taylor and Sarno (1998) developed tests that permit the autoregressive parameters to differ across panel members under the stationary alternative, but because they are joint tests of the null hypothesis, they are not informative about the number of series that are stationary processes when the null hypothesis is rejected. Breuer et al. (2001) further claimed that, by analogy to simple regression, when an F-statistic rejects the null that a vector of coefficients is equal to zero, it does not follow that each coefficient is nonzero. Similarly, when the unit-root null hypothesis is rejected, it may be erroneous to conclude that all series in the panel are stationary. To avoid the problem, Breuer et al. (2001) introduced the "seemingly unrelated regressions augmented Dickey-Fuller"(SURADF) test, which is an augmented Dickey-Fuller test based on the panel estimation method of seemingly unrelated regression (SUR). The system of the ADF equations we estimate here are:

$$\Delta X_{1,t} = \alpha_1 + \beta_1 X_{1,t-1} + \gamma t + \sum_{j=1}^{k1} \theta_{1,j} \Delta X_{1,t-j} + \varepsilon_{1,t} \quad t = 1,2,....,T$$

$$\Delta X_{2,t} = \alpha_2 + \beta_2 X_{2,t-1} + \gamma t + \sum_{j=1}^{k2} \theta_{2,j} \Delta X_{2,t-j} + \varepsilon_{2,t} \quad t = 1,2,....,T$$

. . . .

. . . .

$$\Delta X_{N,t} = \alpha_N + \beta_N X_{N,t-1} + \gamma t + \sum_{j=1}^{kN} \theta_{N,j} \Delta X_{N,t-j} + \varepsilon_{N,t} \quad t = 1,2,....,T \qquad [4]$$

We test the N null and alternative hypotheses individually:

$H_0^1 : \beta_1 = 0; H_A^1 : \beta_1 < 0$

$H_0^2 : \beta_2 = 0; H_A^2 : \beta_2 < 0$

. .

. .

$H_0^N : \beta_N = 0; H_A^N : \beta_N < 0$

with test statistics computed from the SUR estimates of system [4]. As Breuer et al. (2001) showed the imposition of an identical lag structure across panel members could bias test statistics, we select the lag structures for each equation based on the method of Perron (1989).

The major difference between the SURADF and other panel unit tests derives from the formulation of the null hypothesis. While the others are joint tests of a unit root for all members of the panel, the SURADF tests a separate unit-root null hypothesis for each individual panel member and, therefore, identifies how many and which series in the panel are stationary processes.

4. EMPIRICAL RESULTS and POLICY IMPLICATIONS

For comparison, we first apply several conventional unit root tests to examine the null of a unit root in the unemployment rate of each country. We select the lag order of the test on the basis of the recursive t-statistic, as suggested by Perron (1989). The results in Table 2 clearly indicate that the ADF, DF-GLS (of Elliott et al., 1996), the P-P and NP (of Ng and Perron, 2001) tests all fail to reject the null of non-stationary unemployment for all ten countries. The KPSS test also yields the same results. Since the single-equation ADF test has low power with short time spans, as pointed out by Shiller and Perron (1985), here we only have annual observations spanning a 39-year period, perhaps indicating that the failure of the ADF test to have previously rejected the unit root null was due to the time span of the data. We investigate this possibility by exploiting the cross-section variability among regions by applying the Levin-Lin (1992), Im-Pesaran-Shin (1997), Taylor and Sarno (1998) and Breuer et al. (2001) panel-based unit root tests and examine the stationarity of unemployment. Table 3 shows that the panel-based unit root test results are also indicative of the nonstationary unemployment rates. It seems reasonable to conclude that the hysteresis hypothesis as it applies to the unemployment rates for the ten European countries studied cannot be rejected.

Worth noting is that the results here are not consistent with those of Song and Wu (1997, 1998) which, based on the unemployment rate data for 48 U.S. states and 16 OECD countries, support the weak version of the natural-rate hypothesis. Our results, nevertheless, are consistent with those of Leon-Ledesam (2002), which support the notion of hysteresis in unemployment for the European countries.

Table 4 presents Breuer et al.'s (2001) SURADF test results, which indicates the hysteresis hypothesis holds true for all the European countries studied here with the exception of Belgium and the Netherlands.

A major policy implication of our study is that a stabilization policy may have some permanent effects on the unemployment rates of the European countries under study. What, however, are the most effective policies to fight this continuously climbing unemployment? To answer this, the underlying reasons for unemployment must first be identified, but as this is beyond the scope of this paper, it will be investigated in a future study.

5. CONCLUSIONS

In this study, we employ the Levin and Lin (1992), Im et al., (1997) and Taylor and Sarno (1998) and Breuer et al, (2001) tests of the four panel-based unit root types to assess the hysteresis hypothesis in unemployment using data from selected European countries. The results based on the first three panel-based unit root tests indicate that the hysteresis hypothesis as it applies to unemployment cannot be rejected for the European countries studied. Breuer et al’s (2001) SURADF test also indicates the hysteresis hypothesis is supported for all the European countries except for Belgium and the Netherlands.

Finally, as concerns major policy, our study implies that a fiscal stabilization policy would possibly have permanent effects on the unemployment rates of these European countries.

ACKNOWLEDGEMENTS

The authors are grateful to Professors Peter Pedroni and Myles S. Wallace who kindly provided the RATS program codes as well as Professor David E. Rapach who kindly provided the GAUSS program codes, respectively. Any errors that remain are our own.

References

Blanchard, O and Summers, L., 1986, Hysteresis and the European unemployment problem, in *NBER Macroeconomics Annual*, Vol. 1. Cambridge: MIT Press.

Breuer, Janice Boucher, McNown, Robert., and Wallace, Myles S., 2001, Misleading inferences from panel unit-root tests with an illustration from purchasing power parity, *Review of International Economics*, 9(3), 482-493.

Brunello, G., 1990, Hysteresis and 'the Japanese unemployment problem: a preliminary investigation, *Oxford Economic Papers*, 42, 483-500.

Elliott, G., Rothenberg, T J., and Stock, J H., 1996, Efficient tests for an autoregressive unit root, *Econometrica,* 64, 813-836.

Im, KS., Pesaran, M. H., and Shin, Y.,., 1997, Testing for unit roots in heterogeneous panels, *DAE Working Papers 9526*, University of Cambridge.

Jaeger, A. and Parkinson, M., 1994, Some evidence on hysteresis in unemployment rates, *European Economic Review*, 38, 329-342.

Levin, A., Lin, C.F., 1992, Unit root in panel data: Asymptotic and finite-sample properties, Discussion Paper No. 93-56, Department of Economics, University of California, San Diego.

Leon-ledesma, M A., 2002, Unemployment hysteresis in the US states and the EU: a panel approach, *Bulletin of Economic Research*, 54,2, 95-103.

MacDonald, R., 1996, Panel unit root tests and real exchange rates, *Economics Letters*, 50, 7-11.

Maddala,G.S. and Wu, Shaowen., 1997, A comparative study of unit root tests with panel data and a new simple test, Manuscript, Ohio State University.

Mark, N C., 2001, International Macroeconomics and Finance, Chapter 2, Blackwell Publishers Inc., USA.

Mitchell, W F., 1993, Testing for unit roots and persistence in OECD unemployment rates, *Applied Economics*, 25, 1489-1501.

Ng, Serena. And Perron, Pierre., 2001, Lag length selection and the construction of unit root tests with good size and power, Econometrica, 69(6), 1519-1554.

O'Connell, P.G.J., 1998, The overvaluation of purchasing power parity, *Journal of International Economics*, 44, 1-20.

Oh, K-Y., 1996, Purchasing power parity and unit root tests using panel data, *Journal of International Money and Finance*, 15, 3, 405-418.

Papell, D H., 1997, Searching for stationarity: Purchasing power parity under the current float, *Journal of International Economics*, 43, 313-332.

Papell, D H. and Theodoridis, H., 2001, The choice of numeraire currency in panel tests of purchasing power parity, *Journal of Money, Credit and Banking*, 33, 2, 790-803.

Perron, P., 1989, The great crash, the oil price shock and the unit root hypothesis, *Econometrica*, 57(6), 1361-1401.

Rapach, D E., 2002, Are real GDP levels nonstationary? Evidence from panel data tests, *Southern Economic Journal*, 68,3, 473-495.

Roed, K., 1996, Unemployment hysteresis – macro evidence from 16 OECD countries, Empirical Economics, 21, 589-600.

Shiller, R.and Perron, P., 1985, Testing the random walk hypothesis: power versus frequency of observations, *Economics Letters*, 18, 381-386.

Song, F.M. and Wu, Y., 1997, Hysteresis in unemployment: evidence from 48 U.S. states, Economic Inquiry, 35, 235-244.

Song, F.M. and Wu, Y., 1998, Hysteresis in unemployment: evidence from OECD countries, Quarterly Review of Economics and Finance, 38, 181-192.

Taylor, M. and Sarno, L., 1998, The behavior of real exchanges during the post-Bretton Woods period, *Journal of International Economics*, 46, 281-312.

Wu, J-L. and Wu, S., 2001, Is purchasing power parity overvalued? *Journal of Money, Credit and Banking*, 33, 3, 804-812.

Wu, Y., 1996, Are real exchange rates nonstationary? Evidence from a panel-data test, *Journal of Money, Credit and Banking*, 28, 1, 54-63.

Zellner, A., 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association*, 57, 348-368.

Table 1. Summary Statistics of Unemployment Data Sets

Country name	Mean	Std	Maximum	Minimum	Skewness	Kurtosis	J-B
1. Belgium	5.993	3.427	10.811	1.336	-0.097	1.436	4.035
2. Denmark	4.789	3.082	10.486	0.589	-0.082	1.579	3.322
3. France	6.406	3.967	12.399	1.163	0.074	1.438	4.002
4. Ireland	9.915	4.237	16.809	4.978	0.335	1.557	4.116
5. Italy	7.429	2.517	11.837	3.536	0.368	1.899	2.851
6. Netherlands	4.678	3.001	11.693	0.444	0.038	2.025	1.554
7. Portugal	5.022	2.329	8.709	1.619	-0.017	1.448	3.917
8. UK	5.816	3.572	11.396	1.081	0.127	1.516	3.685
9. Norway	2.775	1.491	5.959	1.295	0.899	2.369	5.895*
10. Finland	5.892	4.731	17.031	1.197	1.137	3.092	8.422**

Note: Std denotes standard deviation and J-B denotes the Jarque-Bera Test for Normality. *, **, and *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

Table 2. Univariate Unit Root Tests (ADF, DF-GLS, P-P, KPSS and NP)

Country name	ADF	DF-GLS	P-P	KPSS	NP
1. Belgium	-1.586(1)	-1.243(1)	-1.061[1]	0.617[5]**	-2.521
2. Denmark	-1.218(0)	-0.907(0)	-1.308[3]	0.573[5]**	-1.455
3. France	-1.693(0)	0.097(1)	-1.878[5]	0.725[5]**	0.478
4. Ireland	-1.485(1)	-1.291(1)	-1.299[2]	0.554[5]**	-3.487
5. Italy	-0.398(0)	-0.009(0)	0.389[3]	0.747[5]***	0.252
6. Netherlands	-2.227(1)	-1.115(1)	-2.135[4]	0.562[5]**	-1.455
7. Portugal	-1.665(0)	-1.077(0)	-1.709[1]	0.513[5]**	-1.836
8. UK	-1.301(2)	-0.369(2)	-1.881[9]	0.636[5]**	-0.428
9. Norway	-1.304(0)	-1.071(0)	-1.204[5]	0.621[5]**	-2.343
10. Finland	-1.963(1)	-0.373(2)	-1.256[7]	0.685[5]**	-0.659

Note: *, **, and *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

The number in parenthesis indicates the lag order selected based on the recursive t-statistic, as suggested by Perron (1989). The NP test was based on the MZa statistic.

Table 3. Panel-Based Unit Root Tests

		Critical Values		
A. Levin-Lin ADF Test		0.01	0.05	0.10
-3.648	(k=0)	-8.160	-7.552	-7.264
-5.681	(k=1)	-8.051	-7.473	-7.159
-4.049	(k=2)	-7.888	-7.318	-6.991
-4.363	(k=3)	-7.867	-7.301	-6.994
B. Im-Pesaran-Shin ADF Test				
0.642	(k=0)	-1.489	-1.231	-1.091
-0.114	(k=1)	-1.501	-1.271	-1.112
0.468	(k=2)	-1.492	-1.253	-1.108
-0.881	(k=3)	-1.522	-1.244	-1.177
C. Panel - SUR				
-6.101	(k=0)	-9.403	-8.659	-8.330
-7.611	(k=1)	-9.471	-8.793	-8.385
-5.353	(k=2)	-9.434	-8.514	-8.114
-5.592	(k=3)	-9.565	-8.661	-8.183
D. MADF (Wald Test)				
81.812	(k=0)	125.889	105.214	98.190
126.430	(k=1)	122.498	105.438	97.402
55.995	(k=2)	123.820	104.492	97.184
66.486	(k=3)	124.097	107.034	98.284

Note: *, **, and *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively. Critical values are calculated by Monte Carlo simulation with 10,000 draws, tailored to the present sample size. (For details of this simulation, see Rapach, 2002)

Table 4. SURADF Tests and Critical Values

Country panel label	SURADF	Critical values		
		0.01	0.05	0.10
1. Belgium	-4,312**	-4.465	-4.265	-4.079
2. Denmark	-1.359	-3.346	-2.727	-2.688
3. France	-3.114	-3.367	-3.250	-3.129
4. Ireland	-2.571	-3.209	-2.949	-2.865
5. Italy	-1.814	-4.031	-3.234	-2.812
6. Netherlands	-4.738***	-3.171	-3.067	-3.027
7. Portugal	-2.911	-4.378	-3.399	-3.179
8. UK	-2.112	-3.926	-3.316	-3.192
9. Norway	-1.474	-2.626	-2.496	-2.122
10. Finland	-2.821	-3.664	-3.179	-3.041

Note: *, **, and *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

Critical values are calculated by Monte Carlo simulation with 10,000 draws, tailored to the present sample size. (For details of this simulation, see Breuer et al., 2001)