# METHODOLOGICAL APPENDIX

ONLINE SUPPLEMENT

to article in

JOURNAL OF EUROPEAN PUBLIC POLICY, VOL. 15, ISSUE 8,

2008

## WHO LEADS, WHO FOLLOWS?
## RE-EXAMINING THE PARTY-ELECTORATE LINKAGES ON EUROPEAN INTEGRATION

**Johan Hellström**
**Umeå University, Sweden**

[Update: Additional robustness tests, that uses Chapel Hill expert survey data to measure party positions, are available in Hellström, Johan (2009). Dynamic Interactions : National Political Parties, Voters and European Integration, Umeå University, Department of Political Science, Doctoral thesis. Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-25925]

---

This appendix contains supplementary methodological discussions regarding model specifications for the Granger causality tests, choice of estimation strategies, heteroscedaticity tests, serial correlation tests, multicollinearity diagnostics, unit root tests, and selection of optimal lag length for the Granger causality tests, along with descriptions of the statistical software and data used in the analysis.

### THE GRANGER CAUSALITY TEST

The Granger causality test (Granger, 1969), or the Wiener-Granger method, was originally designed to investigate and determine causal structure in bivariate time series, but recently similar approaches to the

systematic testing and determination of causal directions in panel data settings have been proposed (Arellano & Bond, 1991; Arellano & Bover, 1995; Blundell & Bond, 1998; Holtz-Eakin *et al.*, 1988). The Granger causality test has obvious limitations since it is conditioned on the set of variables included (or omitted), the size of the sample, and the number of lags used (see for example Holtz-Eakin *et al.*, 1988). Therefore, the test should not be regarded 'definitive', but rather as contingent on the choice of variables and lags introduced.
In this study two separate sets of panel autoregressive models (PVARs) are applied to a pooled, unbalanced cross-sectional panel of data pertaining to the stances of political parties and aggregated voter opinion in 15 countries (the 15 EU-members as of 1995) between 1973 and 2003.

To illustrate the Granger causality approach to analysing the empirical causal structure of national political parties' positions and electorates' opinions on European integration consider the within-group (i.e. LSDV)[1] estimations 1 and 2, which combines the traditional VAR approach in a panel data setting, controlling for unobserved heterogeneity (estimates obtained from these models are presented in column 2 of tables 1 and 2 in the article):

$$P_{it} = \alpha_{1,i} + \sum_{j=1}^{n} \beta_{1,j} P_{i,t-j} + \sum_{j=1}^{n} \rho_{1,j} V_{i,t-j} + \eta_{1,i} + \gamma_{1,i} + \varepsilon_{1,it} \quad (1)$$

$$V_{it} = \alpha_{2,i} + \sum_{j=1}^{n} \beta_{2,j} V_{i,t-i} + \sum_{j=1}^{n} \rho_{2,j} P_{i,t-j} + \eta_{2,i} + \gamma_{2,i} + \varepsilon_{2,it} \quad (2)$$

In the above equations, $V$ denotes voter opinion and $P$ party position, $j$ is the lag length (here $j$=1) , $n$ parties in equation 1 or $n$ voter groups (indexed by $i$) are observed over $T$ periods (indexed by $t$), $\eta_i$ represents a set of dummy variables that capture all unobserved group-specific time invariant effects, $\gamma_i$ represents time dummies to account for trending behaviour, and $\varepsilon_{it}$ represents the disturbances, which are usually assumed to be uncorrelated across groups and time (although this assumption is relaxed, as I mention below).

[1] Strictly, I estimate within-group fixed effects models in the article, but for simplicity the Least Square Dummy variable (LSDV) models are shown here, since these two estimators are mathematically equivalent.

Since a single lag is used in the regressions, I estimate party positions using a lag of themselves and on a lag of party supporters' opinions. If the lagged voter opinion variable (i.e. the coefficient $\rho_1$ in equation 1) is significantly different from zero, voter opinions Granger-causes changes in party positions. In the second regression, I estimate voter opinion based on a lag of itself and on a lag of party positions. If the lagged party position variable (i.e. the coefficient $\rho_2$ in equation 2) is significantly different from zero, parties' positions Granger-cause changes in party supporters' opinions.[2]

In my dataset, the number of time periods available for each of $i^{th}$ group, $T_i$, is small and the number of groups, *n*, is relatively large. Since standard panel data estimators yield biased and inconsistent estimates (i.e. the so called Hurwicz or Nickell bias) for short dynamic panels (Nickell, 1981),[3] this implies that the statistics associated with Granger causality tests are not reliable when $T_i$ is small. Therefore, to avoid problems of bias and inconsistency in estimators originating from the inclusion of lagged dependent variables as in equations (1) and (2), I use Difference Generalized Method of Moments (GMM) and System GMM dynamic panel estimators developed by Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998); these are designed for dynamic "small-T, large-N" panels. More specifically, I follow Blundell and Bond (1998) and use the GMM system estimator with forward orthogonal deviations, an alternative to differencing that preserves the sample size in panels with gaps and reduces potential biases in finite samples associated with the difference estimator.

Time dummies were used in all estimated models, since the autocorrelation tests and the robust estimates of the coefficient standard errors assume no correlation across individuals in the idiosyncratic disturbances (see Roodman, 2006). The lagged dependent variable is specified as endogenous to the dependent variable in the estimated

---

[2] In other words, I test the null hypothesis that shifts in voter opinions do not "Granger cause" changes in party positions $(\rho_1 = 0)$ and the null hypothesis that party positions do not "Granger cause" shifts in voter opinions $(\rho_2 = 0)$.

[3] This is a result of there being a correlation between the error terms ($\varepsilon_{1,it}$ and $\varepsilon_{2,it}$) and the lagged dependent variable(s).

models. In the estimated models in tables 3 and 4 all control variables were placed in the instrument's matrix and treated as exogenous.

All singleton observations (i.e. cases where there was only one observation per panel) were removed from all models and the different models were estimated using the same subset of the data (that is, the remaining dataset with two or more time-series observations).

**PANEL DATA DIAGNOSTICS**

A number of statistical issues are typically associated with panel data, namely panel (i.e. groupwise) heteroscedasticity, serial correlation, and multicollinearity (Beck, 2001; Plumper *et al.*, 2005). With respect to groupwise heteroscedasticity, the modified Wald statistic suggested by Greene (2000:598) indicates the presence of panel heteroscedasticity. Serial correlation, or temporal dependence, may also be present in this context because a party's stance on European integration at time $t_{-1}$ may influence its position on European integration at time $t$ (cf. Hellström, 2008:198-199). A standard way to model the dynamics associated with panel data is to include a lagged dependent variable as a right-hand side variable (Keele & Kelly, 2006). Since the Granger causality test already includes lagged dependent variables by default, it should account for such dynamics in the data without further manipulation.[4] Nonetheless, various serial correlation tests applied to asses remaining serial correlation in the estimated models (see table 1 below) indicate that most estimated models contain first order autocorrelation, i.e. AR(1). Nevertheless, when investigating the residuals and estimating the autocorrelation coefficient of these estimations (not shown here), it is evident that the remaining serial correlation is fairly small. Therefore, in these estimations I use heteroscedasticity- and autocorrelation-consistent (HAC) standard errors. In tables 1-4, for the pooled OLS models, within-group and first difference models, I use Bartlett kernel function (also called Newey-West) robust standard errors corrected for autocorrelation and heteroscedasticity (Newey & West, 1987). [5]

[4] When the time-series is stationary and follows a first order serial correlation.

[5] I obtained analogous results when using linear regression with Huber-White standard errors (clustering on party); panel corrected standard errors; and no error correction at all. Correspondingly, for the fixed effects models I get similar results using Arellano robust standard errors; Huber-White standard errors (clustering on party); and no error

*Table 1* Serial correlation tests

| | Table 1 | Table 2 | Table 3 | Table 4 |
|---|---|---|---|---|
| Wooldridge | F(1,78)=1.55<br>Pr.>F=0.217 | F(1,73)=10.84<br>Pr.>F=0.001 | F(1,73)=15.45<br>Pr.>F=0.000 | F(1,70)=9.33<br>Pr.>F=0.003 |
| Arellano-Bond test for: | | | | |
| AR(1) | z = -2.28<br>Pr.>z= 0.022 | z = -2.07<br>Pr.>z= 0.038 | z = -1.34<br>Pr.>z= 0.180 | z= -1.82<br>Pr.>z= 0.069 |
| AR(2) | z = 0.49<br>Pr.>z= 0.622 | z= 0.49<br>Pr.>z= 0.627 | z = 0.17<br>Pr.>z= 0.866 | z = 0.39<br>Pr.>z= 0.696 |

Note: The first test is the Wooldridge (2002:282-83) test for autocorrelation in panel data; the second test is the Arellano-Bond (1991) test for autocorrelation. Both tests have a null hypothesis of no serial correlation.[6]

The Bartlett kernel-based estimator for the standard errors was also used to account for arbitrary patterns of autocorrelation and heteroscedasticity in the Difference and System GMM dynamic panel regression. In addition, in the two-step system GMM, Windmeijer finite-sample correction of standard errors where used (Roodman, 2006). Nonetheless, the consistency of the GMM estimator depends on the validity of the assumption of no serial correlation in the error term (this approach assumes no second-order autocorrelation in the first-differenced idiosyncratic errors.) [7] and on the validity of the instruments. To check for serial correlation and that the instruments are correctly specified I perform two tests: the Arelleno-Bond test for second-order serial correlation of the differenced residuals, and the Hansen test of over-identifying restrictions (as the more common Sargan test is inconsistent because I use the robust standard errors, for further details see below). Full details regarding these tests and the estimation

correction at all. That is, the results are not sensitive on the error correction used to account for panel heteroskedasticity and serial correlation.

6 These tests were performed using the Stata programs 'xtserial' (Drukker, 2003) and 'abar' (Roodman, 2006).

7 In other words, the GMM estimator is consistent if there is no second-order serial correlation in the error term of the first-differenced equation, that is, it requires that $E[\Delta u_{it}\Delta u_{it-2}]=0$. However, this test is only defined for $T\geq 5$, since it involves differenced residuals two periods apart.

procedure can be found in Arellano & Bover (1995) and Wawro (2002). These tests can be found in tables 1-4 in the article. Since the AR(2) tests are not significant, the Arellano–Bond autocorrelation tests in the GMM models indicate that there are no problems relating to serial correlation in levels. The Hansen test for the validity of over-identifying restrictions and the quality of instruments is implemented for each regression and failure to reject the null hypothesis indicates that the instruments are valid, thus supporting the validity of the model specification.

Moreover, if multicollinearity is present the parameter estimates will be consistent (except when there is perfect or almost perfect collinearity), but the standard errors of the estimates will be inflated, thus making statistical significance of the estimates less likely. The possible problem of multicollinearity, i.e. when regressors are correlated, can be addressed by variance inflation factors (VIF).[8] The regressors show a mean VIF of 1.11 to 6.74 for the party equations, and between 1.13 and 5.34 for the voter equations. The reason for the high VIFs is that, in equations 1 and 2, I use ideology of parties and voters, respectively, together their square (yielding individual VIFs of between 20 and 28). Therefore, since these covariates measuring ideology are mainly correlated to their squares, and not with other variables, collinearity should not constitute any problem in the estimated models (since the variables of interest in the analysis do not show any sign of high collinearity).

**UNIT ROOTS TESTS**

Stationarity does not seem to present a problem in these data. The coefficient of the lagged party positions and voter opinion, regardless of the variables included and the model estimated, is never close to one for the lagged party positions or for lagged voter opinions (with the exception of the Pooled OLS models). In addition, I tested for stationarity using Fisher's tests for unbalanced panels as developed by Maddala and Wu (1999); these tests rejected the null hypothesis that all

[8] VIF is the factor by which the sampling variance of an estimator is increased as a result of collinearity. $VIF = \frac{1}{1-R_j^2}$, where $R_j^2$ is the square of the multiple correlation coefficient of $X_j$ regressed on the other variables. As a 'rule of thumb', if any VIF value are larger than 10 or the means of all the VIFs are a great deal larger than 1, than there will be a problem with collinearity. See for example Gujarati (2003:348-354; 362-363).

panel series are non-stationary in levels for all party and party supporter variables.[9] Table *2* shows the results of this testing procedure.

*Table 2* Unit roots tests

| Dependent variable | Augmented Dickey-Fuller | Phillips-Perron | Conclusion |
|---|---|---|---|
| Table 1: Voter opinions (national aggregate) | $\chi2(128)=155.51$ Prob>$\chi2=0.049$ | $\chi2(160)=194.99$ Prob>$\chi2=0.031$ | Reject a unit root |
| Table 2: Party positions (national aggregate) | $\chi2(214)=478.91$ Prob>$\chi2=$ 0.000 | $\chi2(232)=487.12$ Prob>$\chi2=$ 0.000 | Reject a unit root |
| Table 3: Voter opinions (party supporter aggregate) | $\chi2(122)=137.63$ Prob>$\chi2=0.158$ | $\chi2(148)=247.34$ Prob>$\chi2=0.000$ | Reject a unit root |
| Table 4: Party positions (party supporter aggregate) | $\chi2(214)=374.02$ Prob>$\chi2=$ 0.000 | $\chi2(232)=478.91$ Prob>$\chi2=0.000$ | Reject a unit root |

Note: The reported statistics are obtained using 1 lagged first difference terms (ADF-test) and 1 periods of serial correlation (PP-test). ADF and PP test the null hypothesis of existence of unit root.

The Phillips-Perron (pp) t-test is not sensitive to the number of lags in the autocorrelation function and has greater power than the Augmented Dickey-Fuller (ADF) test for small samples. Therefore, although ADF and PP give different results for voter opinions (aggregated at non-national level), I conclude that there is no unit root.[10]

## SELECTION OF LAG LENGTHS

Granger causality tests are sensitive to the choice of lag lengths, and these can significantly influence the results. Unfortunately, no sigle method for choosing the lag length is ideal in all cases. Here, the Akaike

[9] As implemented by the Stata command 'xtfisher'.

[10] As the Difference GMM yields similar results to the System GMM estimations in table 3 this would indicate that no unit roots are present.

Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to identify an appropriate lag length (through assessing model fit), resulting in single lags for both the party and the voter equations.

*Table 3* Lag lengths selection

| | Table 1 | | Table 2 | | Table 3 | | Table 4 | |
|---|---|---|---|---|---|---|---|---|
| Lag | ΔAIC | ΔBIC | ΔAIC | ΔBIC | ΔAIC | ΔBIC | ΔAIC | ΔBIC |
| 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 2 | 403.32 | 405.81 | 174.13 | 177.21 | 344.08 | 343.50 | 383.07 | 378.77 |
| 3 | 601.95 | 606.69 | 365.75 | 371.11 | 540.73 | 539.46 | 631.62 | 627.15 |

Note: ΔAIC and ΔBIC denote the AIC and BIC differences, respectively, i.e. the difference between the individual AIC/BIC value and the smallest AIC/BIC over the models reported in tables 1-4.

In addition, I evaluated an alternative strategy for selecting the lag length involving initially testing a long lag and incrementally reducing it, conducting sequential *F* tests on the different lag lengths and stopping when the test rejected the hypothesis that the coefficients are jointly zero. This approach also resulted in single lags for both the party and the voter equations (not shown here).

**STATISTICAL SOFTWARE**

All models were estimated using Stata version 9.2. The pooled OLS models were estimated using the official Stata command 'newey'. For the fixed effects and first difference models the Stata command 'xtivreg2' developed by Schaffer and Stillman (2007) was used. Since this command removes the constant from the fixed effects estimation, this was retrieved from the official 'xtreg' Stata command. The difference and system GMM models were estimated using the Stata command 'xtabond2' developed by Roodman (2006).

**DATA**

All data used in the analysis are described, and their data sources are listed in table 4.

*Table 4* List of variables

| Variables | Mean | Std | Min | Max | Source |
|---|---|---|---|---|---|
| Voter opinions[a] | 0.648 | 0.212 | 0 | 1 | Schmitt *et al* (2005); Eurobarometer 60. |
| Voter opinions[b] | 0.743 | 0.199 | 0 | 1 | |
| Party positions | 0.462 | 0.081 | 0 | 1 | Budge *et al* (2001); Klingemann *et al* (2007). |
| *Control variables* | | | | | |
| Time dummies | | | 0 | 1 | |
| Median age | 42.338 | 2.791 | 32.115 | 50.405 | |
| Proportion female | 0.503 | 0.046 | 0.310 | 0.610 | |
| Manual labourers | 0.155 | 0.055 | 0.040 | 0.360 | |
| Non-manual labourers | 0.181 | 0.062 | 0.055 | 0.458 | |
| Agriculture | 0.019 | 0.023 | 0 | 0.175 | |
| Executives | 0.065 | 0.050 | 0 | 0.280 | |
| Professionals | 0.026 | 0.018 | 0 | 0.132 | |
| Unemployed | 0.055 | 0.027 | 0 | 0.184 | |
| 1st education quartile | 2.034 | 1.218 | 1 | 7 | |
| 2nd educ. quartile | 4.089 | 1.859 | 1 | 9 | |
| 3rd educ. quartile | 7.217 | 1.672 | 1.5 | 10 | |
| Median income | 10.705 | 16.046 | 3 | 96 | |
| Left/right position | 4.218 | 1.735 | 0.307 | 8.989 | Franzmann and Kaiser (2006). |
| Left/right square | 20.794 | 15.270 | 0.094 | 80.808 | |
| Government participation | 0.375 | 0.484 | 0 | 1 | Mackie and Rose (1991; 1998); Woldendorf, *et al* (2000); Koole and Katz (2000; 2001); Katz and Koole (2002); Katz (2003); van Biezen *et al* (2004). |
| Electoral support | 0.146 | 0.133 | 0 | 0.544 | Budge et al (2001); Klingemann *et al* (2007). |

Notes: a.) Congruence in aggregate opinions at the national level b.) Congruence in aggregate opinions at the party level. All control variables for the voter opinions originate from Schmitt *et al* (2005) and Eurobarometer 60.

**REFERENCES**


Arellano, M., & Bond, S. (1991). Some tests of specification for panel data - Monte-Carlo evidence and an application to employment equations. *Review of Economic Studies, 58*(2), 277-297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal Of Econometrics, 68*, 29-51.

Beck, N. (2001). Time-series-cross-section-data: What have we learned in the past few years? *Annual Review of Political Science, 4*, 271-293.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal Of Econometrics, 87*, 115-143.

Budge, I., Klingemann, H.-D., Volkens, A.*, et al.* (2001). *Mapping policy preferences: Estimates for parties, electors, and governments 1945-1998*. Oxford: Oxford University Press.

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal, 3*(2), 168–177.

Franzmann, S., & Kaiser, A. (2006). Locating political parties in policy space - A reanalysis of party manifesto data. *Party Politics, 12*(2), 163-188.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*, 424-438.

Greene, W. H. (2000). *Econometric analysis (fourth edition)*. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2003). *Basic Econometrics (fourth edition)*. New York: NY.: McGraw-Hill.

Hellström, J. (2008). Partisan responses to Europe: the role of ideology for national political parties' positions on European integration. *Journal of European Public Policy, 15*(2), 189-207.

Holtz-Eakin, D., Newey, W., & Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica, 56*(6), 1371-1395.

Katz, R. S. (2003). Political data in 2002. *European Journal of Political Research, 42*(7-8), 873-879.

Katz, R. S., & Koole, R. (2002). Political data in 2001. *European Journal of Political Research, 41*(7-8), 885-896.

Keele, L., & Kelly, N. J. (2006). Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political Analysis, 14*(2), 186-205.

Klingemann, H.-D., Volkens, A., Bara, J., *et al.* (2007). *Mapping policy preferences II: Estimates for parties, electors, and governments in Eastern Europe, European Union, and OECD 1990-2003*. Oxford: Oxford University Press.

Koole, R., & Katz, R. S. (2000). Political data in 1999. *European Journal of Political Research, 38*(3-4), 303-312.

Koole, R., & Katz, R. S. (2001). Special issue: Political data yearbook, 2001 - Political data in 2000. *European Journal of Political Research, 40*(3-4), 223-232.

Mackie, T., & Rose, R. (1991). *The international almanac of electoral history*. Washington, D.C.: Congressional Quarterly.

Mackie, T., & Rose, R. (1998). *A decade of election results: Updating the international almanac*. Glasgow: Centre for the study of Public Policy, University of Strathclyde.

Maddala, G. S., & Wu, S. (1999). A comparative study of the Unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics, 61*, 631-652.

Newey, W., & West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*(3), 703-708.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica, 49*(6), 1417-1426.

Plumper, T., Troeger, V. E., & Manow, P. (2005). Panel data analysis in comparative politics: Linking method to theory. *European Journal of Political Research, 44*(2), 327-354.

Roodman, D. (2006). How to do xtabond2: An introduction to "difference" and "system" GMM in Stata, *Working Paper No. 103*: Center for global development.

Schaffer, M. E., & Stillman, S. (2007). xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. from http://ideas.repec.org/c/boc/bocode/s456501.html

Schmitt, H., Scholz, E., Leim, I., *et al.* (2005). Mannheim Eurobarometer trend file, 1970-2002, Data set edition 2.0. Mannheim/Cologne: MZES/ZUMA/ZA.

van Biezen, I., Katz, R. S., Mackerras, M., *et al.* (2004). Political data in 2003. *European Journal of Political Research, 43*(7-8), 919-1175.

Wawro, G. (2002). Estimating dynamic panel data models in political science. *Political Analysis, 10*(1), 25-48.

Woldendorf, J., Keman, H., & Budge, I. (2000). *Party government in 48 democracies (1945-1998)*. Dordrecht: Kluwer.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge: The MIT Press.