ECONOMETRIC SOLUTIONS VS. SUBSTANTIVE RESULTS: A CRUCIAL TRADE-OFF IN THE TIME-SERIES-CROSS-SECTION ANALYSIS*

Federico Podestà

ISER Working Papers
Number 2003-34

Institute for Social and Economic Research

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates the following centres:

- ESRC Research Centre on Micro-social Change. Established in 1989 to identify, explain, model and forecast social change in Britain at the individual and household level, the Centre specialises in research using longitudinal data.

- ESRC UK Longitudinal Centre. This national resource centre was established in October 1999 to promote the use of longitudinal data and to develop a strategy for the future of large-scale longitudinal surveys. It was responsible for the British Household Panel Survey (BHPS) and for the ESRC's interest in the National Child Development Study and the 1970 British Cohort Study

- European Centre for Analysis in the Social Sciences. ECASS is an interdisciplinary research centre which hosts major research programmes and helps researchers from the EU gain access to longitudinal data and cross-national datasets from all over Europe.

The British Household Panel Survey is one of the main instruments for measuring social change in Britain. The BHPS comprises a nationally representative sample of around 5,500 households and over 10,000 individuals who are reinterviewed each year. The questionnaire includes a constant core of items accompanied by a variable component in order to provide for the collection of initial conditions data and to allow for the subsequent inclusion of emerging research and policy concerns.

Among the main projects in ISER's research programme are: the labour market and the division of domestic responsibilities; changes in families and households; modelling households' labour force behaviour; wealth, well-being and socio-economic structure; resource distribution in the household; and modelling techniques and survey methodology.

BHPS data provide the academic community, policymakers and private sector with a unique national resource and allow for comparative research with similar studies in Europe, the United States and Canada.

BHPS data are available from the Data Archive at the University of Essex
http://www.data-archive.ac.uk

Further information about the BHPS and other longitudinal surveys can be obtained by telephoning +44 (0) 1206 873543.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Acknowledgement:

Prepared for the ECPR Conference - Marburg, 18 - 21 September 2003. This paper is based on work carried out during a visit to the European Centre for Analysis in the Social Sciences (ECASS) at the Institute for Social and Economic Research, University of Essex supported by the Access Research Infrastructure action under the EU Improving Human Potential Programme. I would like to thank Bernhard Kittel and Nathaniel Beck for helpful discussions on various aspects of this paper. All errors are solely my responsibility

Readers wishing to cite this document are asked to use the following form of words:

> **Podestà, Federico (2003) 'Econometric Solutions vs. Substantive Results: A crucial Trade-off of Time-Series-Cross-Section Analysis', '*Working Papers of the Institute for Social and Economic Research*, paper 2003-x. Colchester: University of Essex.**

For an on-line version of this working paper and others in the series, please visit the Institute's website at: http://www.iser.essex.ac.uk/pubs/workpaps/

Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ UK
Telephone: +44 (0) 1206 872957
Fax: +44 (0) 1206 873151
E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

The increasing and massive use of time-series-cross-section data has given the impression that the classical problems of the quantitative comparative political economy (i.e., small N) have been overcome. Nevertheless, space and time combination in the econometric analysis has involved several new methodological problems. Among these problems, a trade-off between the substantive and the technical specification of the model arises. This is because the solutions found in order to solve econometric problems (error term complications, multicollinearity of data and non-stationarity issue) require a re-specification of the model developed in theoretical terms. In order to spread light on this trade-off, I have examined the nexus between social spending and its principal explanatory factors. Specifically, I have applied four leading specification strategies to a common TSCS cross section data set. The specification strategies I have compared have been developed by Garrett and Mitchell (2001), Huber and Stephens (2001), Kittel et al. (2001; 2002), and Iversen and Cusack (2000).

Currently, many students of comparative political economy use pooled time series cross-section (TSCS) analysis in order to explore the relationships between domestic institutions and political arrangements, on the one hand, and economic performances and public policies, on the other hand. Literature on the welfare state development is a typical example of this. An impressive number of pooled TSCS models have been estimated to measure the impact of several socio-economic factors and, above all, partisan power on social expenditure (Hicks and Swank 1992; Hicks and Misra 1993; Huber et al. 1993; Schmidt 1997; Swank 2001; Garrett 1998; Garrett and Mitchell 2001; Iversen and Cusack 2000; Huber and Stephens 2001; Kittel and Obinger 2002).

Although most of these studies use the same dependent variable and similar independent variables (see Kittel and Obinger 2002: 12, for a detailed overview), they test their hypotheses via different model specifications, namely static vs. dynamic model, fixed effect specification vs. no-fixed effect specification, model in levels vs. in changes etc. Scholars choose their favoured model specification according to statistical and/or theoretical arguments. Consequently, the model specification can be considered as a tool for dealing with both technical and substantive problems. Nevertheless, a specific model specification generally is not able to address several statistical and theoretical issues simultaneously. Often a TSCS model may be adequate from a theoretical point of view, but inappropriate in econometric terms. In addition, frequently, we are not able to evaluate this because many published analyses fail to test for, and correct, the main TSCS statistical complications (contemporaneous and auto- correlation of residuals, heteroskedasticity, nonstationarity, and multicollinearity). In other cases, developing statistically appropriate specifications involves the inability of solving theoretically important research questions about the welfare state dynamics. This means that crucial trade-offs may affect model specification building.

Accordingly, the aim of this paper is to compare different specification strategies that have recently come out in welfare state literature. My comparison should allow us to highlight the econometric and theoretical (dis)advantages that these specification strategies involve. The specification strategies here selected have been adopted by some leading scholars of welfare state. They are: a) Huber and Stephens (2001), b) Garrett and Mitchell (2001), c) Kittel and Obinger (2002) and Kittel and Winner (2002), and d) Iversen and Cusack (2000).1 In order to evaluate these specification strategies I have

[1] Although Kittel and Winner's (2002) work constitutes for itself a re-analysis of Garrett and Mitchell (2001) study; I have considered both of these TSCS analyse because Kittel and Winner (2002: 6) focus exclusively on the statistical modelling. By contrast, my paper evaluates specification strategies from an econometric point of

applied them on a common data set. This should also allow us to evidence how TCSC estimates are affected by instability. This is because applying different model specifications on the same data set involves substantively different results.

The remainder of the paper is organized as follows. In Section 1 I summarize the main statistical TSCS complications and presents different model specifications as remedies for these complications. This Section also focuses on the potential trade-offs between the technical and theoretical specification of the model. Sections 2 contrasts Huber and Stephens (2001) strategy with Garrett and Mitchell (2001) strategy. Section 3 and 4 are respectively dedicated to Kittel and Obinger (2002) and Kittel and Winner (2002) specification strategy and Iversen and Cusack (2000) specification strategy. Section 5 draws some conclusions.

1. Model specifications as remedies for statistical and theoretical problems

The impressive popularity of the pooled TSCS analysis is based on the advantages this technique ensues. By repeating cross-sectional observations over time, quantitative political economists are now able to solve traditional problems of comparative research and gain important abilities in terms of data modelling. Firstly, the classical "small N" problem has been overcame by using data sets of $N \times T$ observations, obtained by arraying cross-sectional data on N country-units and T time periods. Secondly, by combining time series and cross-sectional data, capturing space and time variation simultaneously (i.e., short- and long-run effect) is more than a concrete possibility.

Although the TSCS design adoption ensures these advantages, it involves at the same time numerous statistical problems which cause a degrading estimator precision. Most of these complications have to do with the violation of standard OLS assumptions about the error process. Specifically, residuals for regression equations estimated from pooled TSCS data using OLS procedure often allow for temporally and spatially correlation as well for heteroskedasticity. Hicks (1994: 171-72) has brilliantly illustrated the nature and the causes of these problems:

---

view as well as theoretical criteria. As it has been argued a statistically adequate model may involve relevant theoretical drawback. Moreover, while the Kittel and Winner re-analysis is about the Garrett and Mitchell's (2001) total spending model, my dependent variable is social spending (cf. below).

Autocorrelation: errors tend to be no independent from a period to the next (serially correlated) because observations and traits that characterize them tend to be interdependent across time. For example, temporally successive values of many national traits (i.e., population size) tend not to be independent over time.

Contemporaneously correlation: errors tend to be correlated across nations because observations of countries belonging to the same area, such as Scandinavia, may be likely linked together.

Heteroskedasticity: errors tend to be, such that they may have differing variances across ranges or sub sets of nations, because countries with higher values on variables tend to have less restricted and, hence, higher variances on them.

Unit and period effects in the error term: even if we start with data that were homoskedastic and not auto-correlated, we risk to register these complications if our model is misspecified because of omitting factors which capture the differences over time and across space.[2]

In addition to the violation of standard OLS assumptions, we have to take into account two other important sources of imprecise estimates. They are:

Nonstationarity: since many time series - usually utilized by comparative political economists - tend to exhibit a high persistence over time, nonstationarity complication becomes a crucial problem for many empirical studies. This appears more evident if we consider that policy outcomes are virtually by definition integrated. It is well known, for example, that budgets are not cooked up from scratch. Instead, policy-makers typically begin with the last period's budget and make incremental changes as deemed necessary" (Durr 1993: 215).

Multicollinearity: although TSCS data give less collinearity among the variables than pure time series data (Baltagi 2001: 6), multicollinearity is a serious problem for many pooled models. It is not uncommon for aggregate variables ganging the levels of economic, social or political activity of a nation to display substantial time trend. When several trending variables are placed in a multivariate regression model, it is not unusual to find not trivial levels of multicollinearity (Janoski and Isaac 1994: 52).

[2] Hicks (1994: 172) mentions also a fifth error complication. More specifically, errors might be nonrandom across spatial and/or temporal units because parameters are heterogeneous across subsets of units.

The imprecise estimation that all these problems involve has generated an intense debate concerning specific remedies to use for dealing with each of them. The debate has essentially interested the choice of a correct estimation procedure for pooled TSCS regression parameters and the selection of an appropriate model specification. In other words, the found remedies for above mentioned complications have to do with estimation and specification issues.

In the political science literature, the estimation dispute has been well solved by Beck and Katz (1995) proposal. It prescribes to retain OLS parameter estimators, but replace OLS standard errors with panel-corrected standard errors (PCSEs) that take into account the contemporaneous correlation of the errors and perforce heteroskedasticity.3 The great popularity known by this approach has demonstrated that the estimation issue and the found solution for statistical problems, such as contemporaneous correlation of errors and heteroskedasticity, have been overcame. However, as Beck and Katz (1995: 946) argue, any autocorrelation of the residuals must be tested and, then, eliminated before OLS/PCSEs are calculated. Therefore, it is priority considering the developed remedies for autocorrelation of the residuals. These remedies - like those developed for the remaining sources of imprecise estimation - have to do with the specification model issue. Accordingly, when we analyse TSCS data, autocorrelation of residuals, nonstationarity, unit and period effects in the error term, and multicollinearity have to be addressed via technical specification developing.

The model (re)specifications recommended for dealing with the above discussed statistical problems are the following:

Autocorrelation remedies: applied studies deal usually with the autocorrelation of residuals with two alternative approaches suggested by Beck and Katz (1995; 1996) in two different articles. The first approach treats serial correlation as nuisance and correct for it by estimating a model for autoregressiveness, AR(1). By contrast, the second approach treats over-time persistence

---

Nevertheless, since this complication bring us out from a model equation where a single slope coefficient is constant across units and time points, is not addressed in this paper

[3] Beck and Katz (1995) proposal is based on the argument that the generalized least squares approach of Parks produces standard errors that lead to extreme overconfidence, often underestimating variability by 50% or more. Conversely, although it is well known that OLS estimates of TSCS model parameters may not be optimal, they often perform well in practical research situations. However, because the OLS estimates of the standard errors may be highly inaccurate in such situations, it is necessary to estimate them throughout a more robust procedure such as White approach Hence OLS/PCSEs solution.

in the data as a substantive information that should be estimated by including a lagged dependent variable in the set of independent variables. Accordingly, the coefficient of the lagged dependent variable yields an explicit estimate of the extent of stickiness or persistence in the dependent variable. Clearly, the choice between the two approaches involves the necessity of solving a specification dilemma: static specification (AR(1) correction) vs. dynamic specification (lagged dependent variable inclusion).

onstationarity remedy: differencing data is the most obvious approach utilized for cooping with nonstationarity time series. Estimating a model with the dependent variable measured as changes from year to year rather than absolute levels is that it is more likely to be able to pass the tests for unit roots. Obviously, this involves a re-specification of the model from levels to changes.

Remedy for unit and period effects in the error term: the problem of unit and time effects is routinely overcame by estimating a fixed effect model which allows intercept term to vary in order to capture the differences in behaviour over time and space (Judge et al. 1985: 519). The reasoning underlying this solution is the following: while error complications may be caused by model misspecifications, OLS are optimal if the model is appropriately specified. Accordingly, if we have failed to take into account relevant explanatory variables that do not change over time and/or others that do not change across units (Kmenta 1990), including year and country dummy on the right-hand side of the equation should allow us a better modelling of error term and, consequently, less imprecise estimates.

Multicollinearity remedy: variable selection is commonly abused by analysts for coping with multicollinearity data. When two or more independent variables are extremely inter-correlated, then the least interesting variable from a theoretical point of view can be omitted. Here the aim is to reduce the regressors in the model to a less highly correlated set (Fox 1991: 15).

A glance at this list reveals that TSCS analysts have to make several choices in order to develop an econometrically adequate model specification. They have to choose between a static or a dynamic model, a model in levels or in changes, and a fixed effect specification or a no-fixed effect specification to deal with the above discussed statistical problems. Often, combining these options does not involve serious problems. Econometrically, there are no obstacles to opt for a dynamic (rather than a static) model when we work with data in levels or in changes. Nevertheless, combining other remedies may involve relevant trade-offs. The dilemma between a parsimonious model (as a multicollinearity remedy) and a fixed effect specification (as a remedy for unit and period effects in the error term) is a typical example of this. On the one side, the combination of dummy variables may be highly correlated with other independent variables, enhancing multicollinearity problems within the model and reducing the efficiency of the coefficient estimates. On the other side, the exclusion of dummy variables to limit the multicollinearity problem should lead to bias in the coefficient estimators because in specifying the regression model we have failed to include relevant explanatory variables (Avelino et al. 2002: 13-4). Such a trade-off demonstrates that one might risk to obtain inevitably imprecise estimates opting for one remedy rather than the other one.

Obviously, the efficacy of remedies above discussed as well as the arising of potential trade-offs in the specification model developing depend on data we are analysing. Consequently, testing for (non)stationarity, multicollinearity, and uncorrelated and homoskedastic residuals allows us to determine whether a model specification is statically appropriate. Accordingly, the specification strategies examined in the successive pages have been subjected to a battery of tests in order to evaluate their technical adequateness.

However, the specification of the model has not to be judged from a statistical point of view only. Choice of a model specification must be also guided by the research questions. When TSCS analysts have to choose between a static or a dynamic model, a model in levels or in changes, and a fixed effect specification or a no-fixed effect specification, they are equally bound by theoretical criteria. Although a pooled TSCS analysis should allow us to capture short- and long-run effects simultaneously, particular statistical complications, such as nonstationarity, might involve a model re-specification that makes us unable to model both of these effects. More precisely, if we have to difference data and, hence, model changes only, any information about the long-run effects could be lost. This could appear as an important gap according to many scholars of welfare state development. As Huber and Stephens (2001: 58) argue, since important social programs mature over a long period of time, the assumption in any analysis of short-term change, particularly annual change, yields a mockery. Moreover , because partisan effects need considerable time to materialize in social expenditure (Garrett and Mitchell 2001: 168; Kittel and Obinger 2002: 41), modeling short-term effects does not allow us to test leading hypotheses about welfare state expansion (such as partisan politics perspective) in a appropriate and comprehensive way. This means that a TSCS design may generate important trade-offs between the technical model specification and the theoretical model specification. Consequently, the specification strategies - examined in the following sections – have been evaluated in substantive terms, as well as in statistical terms.

Finally, opting for a specific model specification rather than another one may involve substantial effects on the direction, magnitude, or statistical significance of the explanatory variables that are more relevant for testing hypotheses. This may, for instance, regard the inclusion of a lagged dependent variable on the right-hand side of the equation rather than opting for AR(1) correction to with the autocorrelation of residuals. As Achen (2000) argues, when an autoregressive term is put, it often acquires a large, statistically significant coefficient, while many or all of the remaining substantive coefficients collapse to implausible small and insignificant values. Consequently, TSCS analysts could be attempted to choose between lagged dependent variable inclusion and AR(1) correction according to the substantive results they respectively involve. As Beck (1991: 68) highlights, researchers often choose specifications because they give them the coefficients they (or their referees or editors) want. Therefore, a glance to the parameter instability across estimated models should allow us to determine which specification strategies are more “convenient” for testing particular hypotheses.

Before discussing the TSCS models I have estimated according to the specification strategies developed by Huber and Stephens (2001), Garrett and Mitchell (2001), Kittel and Obinger (2002) and Kittel and Winner (2002), and Iversen and Cusack (2000), let me spend some words about data and variables I have used. Obviously, a common data set has been here used in order to replicated and compare these specification strategies. The variables included in this data set are described in Table 1 and have been chosen for two reasons. First, they are probably the most commonly employed variables in the quantitative studies on welfare state development. Second, they recur in numerous models that have been estimated by the analysts which have elaborated the specification strategies under discussion. Therefore, even if none of these studies has been de facto replicated, selecting this set of variables should allow us to adequately compare these strategies and evidence some typical problems of using TSCS design in welfare state development analysis.

Table 1 about here

Social security transfers as percentage of GDP (SSTRAN) is a generic indicator of the overall welfare state effort. Obviously, it has been here used as dependent variable. GDP per capita (RGDPC), the percentage of population under 15 and over 64 years (DEPRATIO), imports plus exports as percentage of GDP (TRADE), unemployment rate (UNEM) constitute a set of socio-economic explanatory variables. Conversely, government share of leftist parties (LEFTCUM) and government share of religious parties (RELCUM) constitute two political explanatory variables. Since these two indicators generally represent the core variables of "the partisan politics hypothesis", i.e. the main hypothesis of many studies on welfare state development, a particular attention have been given them from a substantive point of view. More precisely, the discussion about statistical significance and signs of coefficients has been dedicated to these variable only. Clearly, the partisan politics hypothesis expects plus signs for coefficients of party variables. Moreover, the party strength has been measured according to Huber, Ragin and Stephens (1998) approach. Specifically, the religious party and leftist party cabinet share are measured by the cumulative cabinet shares from the first year for which data are available (1945) to the year of the observation in question .As it will result more clear below, this measurement method has been chosen because it is more adapt than a measurement of the partisan cabinet share to the same year of the dependent variable to be employed to different specification of the model.

2. Huber/Stephens and Garrett/Mitchell specification strategy

Both Huber and Stephens (2001) and Garrett and Mitchell opt for a specification in levels. According to Huber and Stephens (2001: 58), three important reasons make inescapable the level measurement of social spending. First, measuring the annual change in the dependent variable is misleading because annual changes of welfare expenditure are strongly determined by economic cycles. Second, data on advanced industrial democracies (collected principally by the OECD), contain errors and these errors are greatly exaggerated when one moves from levels to annual change with the same set of data. Third, some important social policy programs, most notably pensions but also most other programs in which entitlements are based on the duration and/or level of contributions, mature over a long period of time. Thus, expenditure will gradually climb without any additional legislation by the sitting government. This maturation effect makes a mockery of the assumption in any analysis of short-term change, particularly annual change, of uniform leads and lags - that a given independent variable affects the dependent variable with a constant time lag across cases and through time.4 Moreover, Huber and

[4] Indeed , Kittel and Obinger (2002: 25-6) have evidenced important arguments against the Huber and Stephens' reasoning. First, the effect of economic cycles can be controlled for by including relevant variables. Second, it is true that data errors bias the results in an analysis of dynamics because parts of the dynamics are due to measurement errors. However, if the time series character of the pooled data set is taken seriously, this also affects the levels because the part of the coefficient capturing the time dimension is also partly driven by the measurement errors. Finally, the assumption of uniform leads and lags lies at the heart of any pooling endeavour except if the leads and lags are estimated separately for each country and variable – not to mention

Stephens (2001: 59) argue that the independent variables must also be measured in a corresponding fashion of the dependent variable (i.e., as levels).

Therefore, the choice of a specification in level is essentially guided by theoretical and methodological arguments. However, although the Huber/Stephens strategy and the Garrett/Mitchell strategy share a specification in levels, they are differentiated by the adopted remedies to address particular statistical complications. Even if both these authors use PCSE to deal with heteroskedasticity and contemporaneous correlation of residual (Beck and Katz 1995), they opt for alternative solutions in order to deal with the autocorrelation. Garrett and Mitchell (2001) include a lagged dependent variable on the right-hand side of the equation to take in account dynamics, as suggested by Beck and Katz (1996). More specifically, they treat the coefficient of the lagged dependent variable as an explicit estimate of the extent of stickiness in the social spending. By contrast, Huber and Stephens (2001: 62) use correction for first order auto-regressiveness and imposition of a common rho for all cross-sections, as suggested by Beck and Katz (1995). They argue that important theoretical reasons induces them to choosing the AR(1) remedy rather than the lagged dependent variable solution. First, an analysis including lagged dependent variable will have all the drawbacks of annual change analysis listed above. This is because, conceptually (though not mathematically), this turns the analysis into an analysis of annual change, since once one controls for the level in the previous year, the remaining unexplained variation is the change from the previous year to the current year. Second, the lagged dependent variable inappropriately drives the equation. That is, the correlation between the dependent variable and the lagged dependent variable is so high that almost nothing else matters. These authors claim (2001: 62) that a large part if not most of the extremely high coefficient for the lagged dependent variable is spurious. The very strong correlation between yt-1 and yt is produced primarily by the fact that Xt-1 and Xt are very strongly correlated and only secondarily by the more modest causal effect of yt-1 on yt. On the basis of their historical comparative analyses, they highlight that policy ratchet effect and other policy legacy effects are not so strong to think that the coefficient of lagged dependent variable can be close to one. Consequently, they conclude that an equation without the dependent

---

the possibility of variation between particular sub-periods within countries – which implies that we tend to end up, for example, with a set of single-country time-series regressions. Hence, implicitly, Huber and Stephens make such assumptions, too, because the differences in the levels over time enter the parameter estimates as the contribution by the time dimension. Indubitably, these are important points. Nevertheless, it is equally true that a model specification focused on short-term effects only does not allow us to capture the maturation effect mentioned by (Huber and Stephens 2001) (see next section).

variable lagged one year is a better estimation of the causal dynamics than an equation with it since, though the true coefficient for policy legacies is not 0, it is a lot closer to 0 than to .9.

Therefore, choosing alternative solutions for addressing the autocorrelation problem, the two specification strategies demonstrate that a theoretical trade-off exists between the AR(1) correction and the inclusion of a lagged dependent variable. It derives from different perspective about dynamics modelling. If one assumes that dynamics must be capture via lagged dependent variable inclusion, the autocorrelation problem will equally addressed by this variable. Otherwise, if one assumes that a lagged dependent variable is not an appropriate estimator of social policy stickiness, dynamics will be exclusively captured via explanatory variables and AR(1) correction will be use to deal with serial correlation of residuals.5

A second divergence between these specification strategies emerges with respect to the: multicollinearity cure and the unit and period effects modelling. As it is already mentioned, this involves a trade-off between fixed effect specification versus no-fixed effect specification. On the one hand, the Huber/Stephens strategy is more interested in a multicollinearity solving than a inclusion of country and year dummies for capturing fixed effects in the error term. Huber and Stephens (2001: 59) outline that, since having to measure the independent variables in a corresponding fashion of dependent variable (as levels), the consecutive measurements of the independent variables are necessarily correlated to one another and the independent variables are also more highly inter-correlated than they are if they are measured as change variables. Thus, multicollinearity is a larger problem in the levels data than in the change data. Accordingly, these authors followed two precautions to contain this risk. First, both statistical criteria and empirical evidence have been adopted for selecting variables. More precisely, independent variables are selected according to statistical (correlation rate) and theoretical criteria. Obviously, this should increase the tolerance levels to a level at which we no longer observed the inflated regression coefficients and instability of the coefficients across equations for different time periods (Huber and Stephens 2001: 62). Second, Huber, Ragin and Stephens (1993: 733) argue that with fixed effects for countries it is impossible to examine the impact of variables constant, or nearly so,

[5] Beck and Katz (1996) argued that the two alternative approaches are differenced by the autocorrelation perception one exhibits. The first approach treats autocorrelation as nuisance and prescribes to correct for it, while the second approach treats the over-time persistence in the data as a substantive information that should be estimated by including a lagged dependent variable in the set of independent variables. By contrast, the arguments developed respectively by Huber/Stephens and Garrett/Mitchell have to do with substantive assumptions exclusively.

over the time. Thus, when we used this estimation technique, the effect of variables we consider important could not be assessed directly. In other words, since the fixed effects are highly collinear with some of independent variables, it is recommendable to not control for country fixed effects.

On the other hand, Garrett and Mitchell (2001: 163) consider this a mistake. This is because if a regressor varies only little over time, but greatly across countries, and if the inclusion of country dummies has a substantial effect on the direction, magnitude, or statistical significance of the variable, the appropriate response is not to exclude the country dummies. Rather, the analyst should conclude that the relevant variable is part the underlying historical fabric of a country that affects the dependent variable and that is not captured by any of the time and country-varying regressors. When these fixed effects are taken into account, the apparent effect of year-to-year fluctuations in the variable could well be very different than when country dummies are not included. According to Garrett and Mitchell (2001: 164), a similar type of reasoning suggests that TSCS regressions should also include year dummy variables to take into account time specific effects. If all the countries in the system are subject to common shocks in a given year, this should be captured by a series of year dummy variables lest these shocks contaminate the regressors of direct interest. Therefore, the Garrett/Mitchell strategy is not concerned about the estimation imprecision deriving from the multicollinearity problem. Garrett and Mitchell opt for a fixed effect specification according to substantive reasons. Including country and year dummies allow them to take into account the inherent features of different countries and common shocks in particular years that affect the outcomes of interest, but that are not accurately captured by any of the included regressors.

Nevertheless, this reasoning underlies a technical objective equally. Since excluding country and year dummies means that we have failed to include relevant explanatory variables, the error term should not be adequately modelled. Thus, the no-fixed effects specification is de facto a misspecified model. Conversely, since OLS are optimal if the model is appropriately specified, controlling for year and country effects should allow us to eliminate heteroskedasticity and autocorrelation of residuals and, consequently, obtain more precise estimates. Clearly, this arises the above mentioned technical trade-off between fixed effect specification versus no-fixed effect specification. Consequently, contrasting the Huber/Stephens strategy with the Garrett/Mitchell strategy, the trade-off between capturing unit and period effects in the error term and reducing regressors of the model emerges. Only data analysis can however allow us to spread light on this trade-off.

Addressing the two above discussed trade-offs according to alternative criteria, Huber and Stephens, on the one hand, and Garrett and Mitchell, on the other hand, opt for different specification in levels. The Huber and Stephens'(2001: 65) preferred specification is an OLS/PCSE model with AR(1) correction and excluding country and year dummy variables. It is the following form:

$$y_{it} = \sum \beta^{j} x_{it}^{j} + e_{it}$$

$$e_{it} = \rho_{i} e_{it-1} + v_{it}$$

where y is the dependent variable and x is an independent variable. The subscript i and t denote, respectively, while the subscript j refers to the particular independent variable. The β is parameter estimates, while e is an error term. The Equation b denotes a first-order autoregressive model (i.e., AR(1)), where ρ_i is a coefficients of first-order autoregressiveness.

By contrast, the Garrett and Mitchell's (2001) favoured specification is a OLS/PCSE model including a lagged dependent variable and a full batteries of i-1 country dummy variables and t-1 year dummies. 6 It is the following form:

$$y_{it} = \beta_1 y_{it-1} + \sum \beta^{j} x_{it}^{j} + \sum \beta_n C_{in} + \sum \beta_m T_{tm} + e_{it}$$

where y_{it-1} stands for the first lag of the dependent variable, while β_1 refers to its slope coefficient. C and T denote unit and time dummy variables.

Table 2 displays results for these specification regressions that I have obtained for above cited data set. Equation 1 corresponds to the Huber and Stephens' model specification, while Equation 2 corresponds to the Garrett and Mitchell's model specification. Equation 3 is a model including the lagged dependent variable and year dummies, Equation 4 is a model including the lagged dependent variable and country dummies, and Equation 5 is a model including lagged dependent variable only.

[6] Indeed, Garrett and Mitchell compare results of their "favorite specification" with those including neither lagged dependent variable nor any fixed effects dummies (but with an AR(1) correction) in order to evaluate

These last three equations are not considered by these as their favoured. Nevertheless, they have been equally estimated in order to evidence which particular components of the model specifications are more relevant in respect with the above mentioned trade-offs. Tests in the lower block of Table 2 should allow us to evaluate how a specification strategy (i.e. a specific remedy) affects the overcoming of each statistical problem.

Table 2 about here

Substantively, results I have obtained are consistent with those of Huber and Stephens (2001) and Garrett and Mitchell (2001). Adopting the Huber/Stephens specification the partisan politics hypothesis is corroborated. Both LEFTCUM and RELCUM coefficient exhibit positive sign and are statistically significant. Conversely, the Garrett/Mitchell specification involves opposite findings: both of these coefficients turn out to be statistically insignificant, while RELCUM acquires a negative sign. Garrett and Mitchell (2001: 168) argue that the basic reason of this difference is that the effects of partisanship on social spending are historical and take many years to materialize. Thus, controlling for lagged spending levels certainly takes into account some of these historical dynamics. Moreover, the pattern of country dummy coefficients, which one can interpret as the historical "equilibrium" level of spending in a country when the effects of all other regressors are taken into account, can do equally. These conclusions seem to bring to light that Garrett and Mitchell implicitly agree with Huber and Stephens (2001) about dynamics modelling. More precisely, if one includes a lagged dependent variable (i.e., yt-1), revealing the effect produced by lagged explanatory variables, such as cabinet share of leftist and religious parties (i.e., xt-1), the long-run effect of party strength is effectively weakened (compare Equation 1 and 5 in Table 2). Moreover, if one even includes both year and country dummy variables it collapses entirely (compare Equation 1 and 2 in Table 2). Consequently, if one assumes this perspective, a model specification excluding both lagged dependent variable and fixed effects results theoretically more adequate and findingly more "convenient" with respect to the partisan politics hypothesis. Nevertheless, these considerations about an appropriate model specification do not take in to account econometric criteria. The two specification strategies must be equally evaluated from a technical point of view.

---

these variables contaminate the independent variables of interest. However, the model excluding the lagged dependent variable and any fixed effect coincides with the Huber and Stephens' preferred specification.

Results I have obtained suggest us that the statistical trade-off about no-fixed effect specification vs. fixed effect specification may be overcame opting for the Huber and Stephens specification. Let me illustrate why. On the one hand, Equation 1 appears appropriate to avoid the multicollinearity complication, but does not yield an adequate modelling of error term. Tests show a systematic violation of OLS assumptions, but the independent variables here included do not reveal an high inter-correlation (Mean VIF is 1,77 only).7 On the other hand, Equation 2 involves an increase of multicollinearity problem, leaving the error term incorrectly modelled all the same. Heteroskedasticity, contemporaneously correlation and autocorrelation of residuals persist, while Mean VIF dramatically goes up to 10,88. Consequently, if including both country and year dummy variables does not allow us to obtain an adequate modelling of error term, why should we endure an imprecise estimates of regression coefficients induced by multicollinearity problem? Beck (2001: 285) argues that if variables of interest are being lost because of the inclusion of fixed effects, the researcher must weight the gains from including fixed effects against their costs. Since a fixed effect specification does not appear a good solution for solving panel complications, the costs of including dummy variables are here higher than the their gains.8

Once this trade-off has been addressed, we have to determine which particular components of the adopted model specifications are more influent with respect to the statistical complications. Specifically, since the violation of OLS assumptions remains an unsolved problem even if both country and year dummy are included, an examination of Equation 2, 3, and 4 can be enough to identify the causes of multicollinearity.9 Locking at the Mean VIF of these equations, it results that multicollinearity derives essentially from country dummies inclusion rather than year dummy control.(this diagnostics shows that this statistical problem is present in Equation 2 and 4 only). This is because some of the explanatory variables (namely, LEFTCUM, RELCUM and TRADE) tend to vary considerably across country, but little over time. This involves an high inter-correlation with country dummy variables and, hence,

[7] Most analysts rely on the informal rules of thumb applied to the variance inflation factor (VIF) in order to check multicollinearity. The Mean VIF should not considerably be lager that 1. Usually, it is assumed that it must not be greater than 6.

[8] This is true even if F statistics of Wald test suggest us that both country and year dummies must be included (see the lower block of Table 2).

[9] However, some consideration may be done concerning the unhelpfulness of a fixed effect specification for an adequate error term modeling. When year dummy variables are included, any advantages is aimed in terms of solving autocorrelation problem (compare Durbin “m” test results in Equation 1, 3 and 5). Analogously, when year dummy variables are included, heteroskedasticity and cross-correlation of residuals do not decrease (compare Modified Wald test results and Breusch-Pagan LM test in Equation 1, 4 and 5).

multicollinearity problem. Therefore, excluding country fixed effects would be here enough to avoid estimation imprecision deriving from multicollinearity

Regarding the second trade-off, namely lagged dependent variable inclusion vs. AR(1) correction, my findings reveal that both Huber and Stephens remedy and Garrett and Mitchell remedy are not sufficient to eliminate the autocorrelation of residuals. Durbin “m” test shows that neither the AR(1) correction (performed by Huber and Stephens) nor the lagged dependent variable included (performed by Garrett and Mitchell) are adequate to get objective. The test demonstrates that this complication of residuals persists in Equation 1 and 2. This suggests us that, although choosing an econometric remedy could be supported by theoretical arguments (cf. above), it doesn’t necessarily ensue a technical solution of the problem. Even if we are not able to say a lot regarding data modelled respectively by Huber and Stephens (2001) and Garrett and Mitchell (2001), it is important to note that both of them use several theoretical arguments to give explanation for their model specification, but do not check whether autocorrelation of residuals has been eliminated once their favoured remedy has been employed. Although several year ago Beck and Katz (1996: 10) recommended to test for autocorrelation of residuals, it seems that these their followers have understood part of teaching only. They adopt the alternative solutions proposed by (Beck and Katz 1995; 1996) on the basis of different theoretical arguments, but not make an effort to test for the efficacy of their remedies.

3. The Kittel/Winner/Obinger specification strategy

Apart from these considerations, it remains to solve the autocorrelation problem in data here analysed. However, since the coefficient of lagged dependent variable is very high denoting a strong persistence of data (see tab 2, Equation 2, 3, 4, and 5), we might well be faced with a even more severe problem. As Kittel and Winner (2002: 20) and Kittel and Obinger (2002: 24) argue, we have to suspect nonstationarity and more formal unit roots are in order. Analogously, Beck and Katz (1996: 11) claim that an advantage of including a lagged dependent variable on the right-hand side of the equation is that it allows us to explicitly consider issues of unit root TSCS data. This is because, just as for a single time series, TSCS models have a unit roots if the estimated coefficient of lagged dependent variable is one. Therefore, since my results are very close to this situation, formal unit roots tests escapable.

Table 3 about here

Consequently, I have performed Levin-Lin test and Im-Pesaran-Shin test for panel unit roots according to the Kittel/Winner and Kittel/Obinger approach. The findings of various specifications of these tests confirm that social security transfers are affected by the nonstationarity problem (see Table 3). All of the test specifications do not reject the nonstationarity hypothesis.

This means that, if one adopts the Huber/Stephens and/or Garrett/Mitchell specification strategy to analyse these data, a further and, likely, more serious source of estimation imprecision arises. As it is well known, when we are dealing with integrated processes and opt for a simple OLS regression in level, the result is a spurious regression problem. This means that there is no sense in which dependent and independent variables are related, even if a OLS regression using the usual t statistics indicate a relation (Granger and Newbold 1974). 10 Therefore, our findings confirm the Kittel and Obinger (2002: 39) suspect: since social expenditure cannot be regarded as stationary, all analyses using the level as a dependent variable are at risk of leading to spurious inferences.11 Surprisingly, neither Huber and Stephens nor Garrett and Mitchell have seriously dealt with this risk. Although the coefficient of the lagged dependent variable of the social security transfers of the Garrett and Mitchell's model is 0,939 (Garrett and Mitchell 2001: 166), they have not considered the unit root possibility and, hence, spurious regression danger. Indeed, Huber and Stephens have addressed this problem by including a time trend variable in the equation in order to detrend the data.12 Nevertheless, when the processes are integrated of order one (cf. below), time trend inclusion could not be enough (Wooldridge 2002: 584). Consequently, Huber and Stephens have likely adopted an inadequate remedy.

However, nonstationarity of social security transfers measured as levels induce us to test whether differencing data ensue the solution of the problem. Because all specifications of Levin-Lin test and Im-Pesaran-Shin test reject nonstationarity (see Table 4), we have obviously model social security

---

[10] Indeed, I have addressed the (non)stationarity issues for the dependent variable only. No test has been performed regarding independent variables. Nevertheless, the core variables of the model, i.e. government share of left parties and government share of religious democratic parties, being measure in cumulative terms, should exhibit nonstationarity by construction (Kittel and Obinger 2002: 26). Consequently, there good changes to think that the OLS regression in level have been performed are spurious regressions and, more importantly, relationships between partisan variable and social security transfers are misleading.

[11] Kittel and Obinger (2002) have performed these panel unit tests on a data set covering the period 1982-1997. Here, the period under examination is 1961-1991. This means that, while my analysis regards a typical phase of the welfare state development, the Kittel and Obinger's study is about a period of welfare state reorganisation.

transfers as changes and assert that social security transfers are integrated of order one. More precisely, we have to opt for model specifications that include on the left-side hand of the equation a variable measured in changes.

Two different specifications appear here available (Beck and Katz 1996: 11): the first difference model and the error correction model. The first-difference model constitutes the specification strategy adopted by Kittel and Winner (2002), while the error correction model constitutes the specification strategy followed by Iversen and Cusack (2000).13

The Kittel and Winner's (2002: 25-8) favoured specification is an OLS/PCSE first-difference model including both lagged dependent variable and fixed effects. 14 This model has the following form:

$$\Delta y_{it} = \beta_1 \Delta y_{it-1} + \sum \beta_{\Delta}^{j} \Delta x_{it}^{j} + \sum \beta_n C_{in} + \sum \beta_m T_{tm} + e_{it}$$

where C and T denote respectively country and time dummy variables, while Δ is the first difference operator. Table 4 displays results I have obtained by applying this model specification to data above described. Equation 1 is a model including the lagged dependent variable plus year and country dummies, Equation 2 is a model including the lagged dependent variable plus country dummies, Equation 3 is a model including the lagged dependent variable plus year dummies, while Equation 4 is a model including the lagged dependent variable only.

Since differencing data is for itself a remedy to the autocorrelation problem (Janoski and Isaac 1994: 34), eliminating this error term complication does not appear a serious difficulty here. According to the

---

[12] Obviously, this equation does not correspond to the Huber and Stephens' favored model.

[13] The Kittel and Obinger's (2002: 28) favoured model specification is not here applied in order to keep well distinct the first-difference specification from the error correction model. The Kittel/Obinger specification can not be treated neither as a pure first-difference model nor a pure error correction model. Although these researchers are not interested to model the long-run effects via their TSCS analysis, they control for lagged social spending levels as prescribed by error correction specification. Nevertheless, they (2002: 27) admit that this term is included with respect to theoretical reasons essentially. It should allow them to test whether higher levels of social expenditure should have a diminishing impact on its growth, reflecting a slow down of social expenditure growth in the more advanced welfare states as compared to the welfare state laggards.

[14] Since the F-test on the country effects - performed by Kittel and Winner (2002: 24) - indicated that the country effects are insignificant, their favored specification includes time effects only. Nevertheless, this result is not inevitable performing a first difference model, I have estimated all possible variants of a fixed effect specification. This allows us also to evidence which particular components of the model are more relevant in respect with the above mentioned trade-offs. The same criteria has been adopted for Huber/Stephens and Garrett/Mitchell specification strategies.

Durbin “m” test (see the lower block of Table 4), the autocorrelation of residuals has been removed in all estimated models to exception of the Equation 4. Nevertheless, multicollinearity might here be a source of degrading estimator precision. Although this complication is a larger problem in the levels data than in the change data (Huber and Stephens 2001: 59), it comes especially out when year dummies variables are included. In fact, the Mean VIF goes dramatically up huge values (see Equation 1 and 2). Obviously, this is because differenced data – denoting changes from year to year – are highly correlated with year dummy variables. Consequently, the trade-off between a fixed effect specification and a no-fixed effect specification may re-emerge in the first-difference specification. Here, the logic is, however, different from the specification in levels. In that case, the trade-off appeared solvable in favour of a no-fixed effect specification because dummy variables have not been able to address the violation of OLS assumptions. Conversely, controlling for fixed effects could be helpful for eliminating the remaining autocorrelation of residuals of a first-difference model. More precisely, including explanatory variables that do not change over time and/or others that do not change across cross-section units, a better specification of the model is obtained and, consequently, a more adequate modelling of error term. If that is the case, we risk to obtain inevitably imprecise estimates. On the one hand, if one opts for a fixed effect specification in order to get a better error term modelling, a degrading precision derives from multicollinearity data. On the other hand, if one opts for a no-fixed effect model in order to avoid multicollinearity problem, a degrading precision derives from the remaining autocorrelation of residuals (plus the other violations of OLS assumptions). My results show that acceptable estimates can be however obtained. Tests in the lower block of Table 4 (see, in particular, Mean VIF and Durbin “m” test) show that Equation 3 and 4 are not dramatically affected by that trade-off.

However, the first-difference model is not affected by technical problems only. As it is well-known, it focuses on systematic associations between the annual changes in the variables – i.e. the short-term effects. This may involve important problems from a theoretical point of view (Janoski and Isaac 1994: 34). Differencing dependent variable and all independent variables of the model, we could not be able to capture important aspects of the welfare state development. As it has been illustrated, the maturation effect of many social programs can not be modelled in any analysis of short-term changes. Moreover, since partisan politics hypothesis assumes that a considerable time is necessary to materialize partisan effect in social expenditure, this can not be captured via a annual change modelling. Nevertheless, the first-difference model can equally give us important information about welfare state dynamics. More precisely, since parties in government can influence the extent of social expenditure growth during their term in office (rather than the levels of expenditure), this model specification can be helpful to capture this relationship (Winner 2002: 22-3). This is feasible, however, only if the (previous) specifications in levels include the government composition in cumulative measurement. In fact, the first difference of these cumulative variables correspond, by definition, to the strength of leftist or religious parties in government during their term in office. This measurement approach is consistent with the partisan politics argument. It is clear that one would not hypothesize that it was only the partisan cabinet share the year before or the average of the previous few years which should determine the level of expenditure or employment in a given year, but rather the cumulative cabinet share over a long period of time (Huber and Stephens 2001: 60). The levels of public (and, hence, social) expenditure should not be related to the strength of leftist or religious parties in government during their term in office but to the cumulative share of these parties in past terms (Kittel and Winner 2002: 23).15

Results I have obtained show, however, that partisanship produces important effects on changes of social security transfers. Although, none of the coefficients of religious parties exhibits the (expected) positive sign and is statistically significant, estimates for leftist parties strength are consistent with the expectations. More precisely, the coefficients of left cabinet portfolios present positive signs and are statistically significant in Equation 3 and 4 (see Table 4). Moreover, if one takes into account that these two equations are not affected by serious statistical complications (cf. above), we can peacefully assert that leftist parties influence the extent of social expenditure growth during their term in office. This finding is particularly interesting if compared with the results of the specification in levels. If one includes both a lagged dependent variable and fixed effects on right-hand side of a regression equation in levels, the year-to-year variations in partisanship seem not to affect social spending in any clear or consistent way (Garrett and Mitchell 2001: 166-8; cf. above).

4. The Iversen/Cusack specification strategy

However, although modelling short-term effects gives us acceptable estimates and important substantive findings, the removal of all level variation from the data remains a relevant drawback of the

[15] A similar logic could be applied to some of the socio-economic variables. Differencing real GDP, we can estimate the effect of growth of the domestic output on social spending. As Kittel and Obinger (2002: 25) illustrate, this allows us to control the effect of economic cycles. Moreover, since Huber and Stephens (2001:61) argue that the aged act as a political lobby group and, consequently, it should be measured via a cumulative approach, differencing data allows us to capture the effect of dependent population as level in the year of observation on annual change of social spending. Because this is not a principal focus of the paper, I have not modelled this dynamics.

first-difference model. Consequently, an error correction specification appears as an inescapable step. Since simple OLS regressions in levels involve spurious inferences, it seems the only solution to capture long-run effects via a TSCS analysis.16

An error correction model could ,however, induce us to test for cointegration. Kittel and Winner (2002: 22) have opted for a first difference specification rather than exploring the possibility of cointegrating relationships because they argue that that approach requires more theoretical elaboration of the expected long-run association between the potentially cointegrating variables. and the literature on panel cointegration is still in its early stages of development. Nevertheless, this does not prevent to estimate an error correction model. As Beck (1991: 70) claims, error correction model is relevant even if cointegration is not.17 Although Engle and Granger (1987) argue that error correction models are especially applicable in the face of cointegration, the principal strength of these models (namely, their ability to capture both long- and short-term relationships among time series) is unrelated to the question of whether or not a series is stationary on its levels (Durr 1993: 191).

Moreover, opting for an error correction model rather than testing cointegration is here more adequate from a theoretical point of view. The error correction model is asymmetric between variables, with a clear distinction between left- and right-hand side variables; the relationship between cointegrated variables is more symmetric (Beck 1991: 70). Therefore, estimating an error correction model for welfare state development implies that social spending adjusts if it is out of equilibrium with socio-economic structure and/or government partisanship, but socio-economic structures and party power do not adjust to move into equilibrium with welfare expenditure. More precisely, we can assume that a decline in leftist and religious party strength translates into a reduction in social spending, but it is unrealistic to assume that a reduction of social security transfers is followed by a redistribution of government composition. Analogously, we would not expect unemployment to decline if unemployment

[16] Kittel and Obinger (2002: 25-41) use the pooled specification to model the dynamics only and leave the long-term level effects to be explored by the simple cross-sectional analyses. Nevertheless, since a cross-section analysis suffers from the problem of few cases and a large number of theoretically relevant independent variables, a parsimonious model must be used to explain variance in the welfare state development (Kittel and Obinger include four independent variables only).

[17] The Engle and Granger (1987) two-step method is the classical method is the classical approach to the cointegration. It proceeds as follows. In the first step a static cointegranting regression of long-run level variables is estimated. If the residuals of the cointegranting regression exhibit stationarity (that is, I(0)), then time series are said cointegrated and we may proceed with the second step regression. In the second step, changes in dependent variable are regressed on changes in independent variable and the previous period's equilibrium term (i.e., the residuals from the cointegranting regression) to estimate the equilibrium rate and the short-run dynamics.

spending is low. This seems here sensible. Conversely, if we think in cointegration terms, it is too easy to allow equilibrium by adjustment of all variables simultaneously. Allowing for everything to correct back to an equilibrium can be appropriate for economics topics, such as the quantity theory of money, but not for the welfare state development argument.

Therefore, the Iversen and Cusack specification strategy does not appear unjustified from a substantive point of view. Although these researchers are not essentially interested in the government composition impact on welfare expenditure, their model specification is more consistent with the "partisan politics hypothesis" assumptions than either a first-difference model or a cointegration test. The error correction model estimated by Iversen and Cusack (2000) has the following form:

$$\Delta y_{it} = \sum \beta_{\Delta}^{j} \Delta x_{it}^{j} + \beta_1 y_{it-1} + \sum \beta^{j} x_{it-1}^{j} + e_{it}$$

It can be considered as a unpacking of the error correction specification recommended by Beck and Katz (1996: 11), that is:

$$\Delta y_{it} = \sum \beta_{\Delta}^{j} \Delta x_{it}^{j} + \phi(y_{it-1} - \sum \gamma^{j} x_{it-1}^{j}) + e_{it}$$

where ϕ represents the rate at which y_{it} and x_{it} return to their long term equilibrium relationship, while the parameter for a change variable, x_{it}^{j} represents the short-term impact of y_{it} on x_{it}. In terms of Iversen and Cusack (2000), γ is freely estimated. By so doing, one can estimate ϕ in his equation, and γ will be such that it is equal to $\phi\gamma$ in the "Beck/Katz ECM formulation". Accordingly, as Iversen and Cusack (2000: 330) highlight, the parameters of the "unpacking error correction model" must be interpreted in the following terms: the parameter for the lagged dependent level variable (β1) provides an easy check on equilibrium properties. β1 should be between –1 and 0 to ensure that the incremental effects of a shock to any exogenous variable are progressively reduced over time, causing spending to converge to a long-term equilibrium. The parameter for a lagged independent level variable, xt–1, is a measure of the permanent (or lasting) effect of a one-off change in that variable, while the parameter for a change variable, xt, remains a measure of the transitory (or passing) effect of a one-off change in that variable. The long-term permanent effect of an independent variable can be calculated by dividing the parameter for the lagged level of that variable by minus the parameter for the lagged dependent level variable: βj/–β1 (assuming that β1 is between 0 and –1). Therefore, this TSCS specification allows us to model the long-run effect of government partisanship on social spending, as well as the maturation effect of social programs as recommended by Huber and Stephens (2001). More precisely, since this model specification gives us the time path of the social spending in levels (given that: Yt = Yt–1+ ΔYt), the maturation effect is effectively captured.

Results I have obtained by estimating this error correction specification are presented in Table 5. Although Iversen and Cusack (2000: 333) estimated a model including country dummy variables only, I have estimated the possible variants of the (no)fixed effect specification. Accordingly, Equation 1 is a model including both year and country dummies, Equation 2 is a model including country dummies only, Equation 3 is a model including year dummies only, while Equation 4 is a model excluding both year and country dummies. Like for the previous specification strategies, this should allow us to evidence which model components are more adequate (or inadequate) for dealing with the different statistical complications.

None of these equations suffers from the autocorrelations of residuals (see Durbin "m" test in the lower block of Table 5). Moreover, the fact that this statistical complication does not affect Equation 4 (i.e. the model excluding both year and country dummies) reveals that a fixed effect specification is not necessary to get a sufficiently adequate error term modelling.18 The autocorrelation problem can be here eliminated via a basic error correction specification. Furthermore, including year and country dummies involves an increase of multicollinearity (compare Mean VIF across Equation 1, 2, 3 and 4). Nevertheless, while the inclusion of country dummy variables maintains the Mean VIF within tolerable limits (9,65 - see Equation 2), the inclusion of year dummy variables involves dramatically elevated values of this statistics (see Equation 1 and 3). As it has been agued about first-difference model, this is because differenced data included on the right-hand side of the error correction model are highly correlated with year dummy variables. Consequently, a control of time fixed effects must be here excluded to avoid an estimation imprecision deriving from multicollinearity problem.

Conversely, including country dummy variables turns out to be a critical issue with respect to the equilibrium proprieties captured by the lagged dependent variable. Although the coefficient of social

[18] Heteroskedasticity and contemporaneous correlation of residuals are addressed via PCSEs.

security transfers on previous year falls between -1 and 0 in all estimated equations, it is statistically significant in Equation 1 and 2 only, namely when country dummy are included. This means that the tendency of reverting to the equilibrium level of social spending is confirmed only if we model the unobserved country-specific variation via a country-specific intercept. Put differently, since the reasoning underlying the fixed effects model is that specifying the regression model we have failed to include relevant explanatory variables that do not change over time (Kmenta 1990), the "predisposition" of social security transfers to converge to a long-term equilibrium is corroborated on condition that we include country dummy variables as a cover-up of our ignorance. This result brings to light an important substantive meaning: theoretical variables are not sufficient for evidencing the equilibrium property of the welfare state development. Since Iversen and Cusack (2000) have controlled for cross-sectional idiosyncrasies equally, re-estimating their model excluding country dummies would be interesting to test for this.

However, it is important to note that party power does not appear here relevant in explaining welfare expenditure variation. None of the cabinet composition variables registers a statistically significant permanent and/or transitory impact on social security transfers. My results - like those obtained by Iversen and Cusack (2000: 333-5) - corroborate the hypothesis that the level of social spending is not necessarily a contentious partisan issue. This finding appears more interesting if it is compared with the results of the previous model specifications. With respect to the Kittel/Winner specification, we can assert that if a term to make sure that social security transfers and political plus socio-economic factors are "on track" is added, the coefficient of the differenced cumulative leftists cabinet (i.e., the variable denoting the strength of leftist parties in government during their term in office) turns out to be statistically insignificant.19 This could be an effect of the not weak correlation between the differenced cumulative leftists cabinet and the cumulative leftists cabinet (that is 0,50). In other words, adding variables to the first-difference specification, substantial effects on the direction, magnitude, or statistical significance of the original estimates could arise as a consequence of the high inter-correlation among the independent variables. With respect to the specification in level suggested by Huber and Stephens (2001) and Garrett and Mitchell (2001), we can assert that adopting a statistically more adequate specification to capture the long-run effects (such as the error correction model), the impact of government partisanship on social spending collapses.

5. Conclusions

Comparing the specification strategies adopted by Huber and Stephens (2001), Garrett and Mitchell (2001), Kittel and Winner (2002), Kittel and Obinger (2002), and Iversen and Cusack (2000) has confirmed that crucial trade-offs affect the TSCS model developing for welfare state development. These trade-offs arise if one has to choose between different specifications for dealing with several statistical complications, as well as different theoretical puzzles.

However, since my analysis has corroborated the Kittel and Obinger's (2002) suspect concerning the nonstationarity of the welfare expenditure, the trade-offs affecting the specifications in levels are de facto cancelled by the unmeaning inferences that these TSCS models involve. More precisely, since the coefficient of the lagged dependent variable of the dynamics models in levels is close to one and panel unit roots tests confirm that social security transfers are affected by the nonstationarity problem, spurious regression problem is here more than a risk. Therefore, although the Huber/Stephens strategy and the Garrett/Mitchell strategy are based on important theoretical arguments (and, at the same time, contrasted by several technical and theoretical dilemmas), applying their favoured TCS model involves spurious regressions that eliminate any justification for a specification in levels.

This means that the first inescapable step in building an adequate TSCS specification for welfare state development is differencing data. Hence, we can opt for either a first-difference model or a error correction model, as suggested by Beck and Katz (1996: 11). Kittel and Winner (2002) specification strategy recommends the first-difference model, while the Iversen and Cusack (2000) specification strategy prescribes the error correction model. Their empirical application has, however, demonstrated that technical trade-offs may arise even when model specifications are able to ensue meaning estimates. Specifically, the trade-off between a fixed effect specification and a no-fixed effect specification may emerge in the first-difference specification, as well as in the error correction specification. This is the case when fixed effects specification involves multicollinearity problem and provides, at the same time, an appropriate error term modelling (i.e. elimination of the autocorrelation problem). Nevertheless, since multicollinearity problem is essentially provoked by time dummies

[19] Beck (1991: 69) refers to the error correction model as a first difference model plus a term to make sure that dependent and independent variables stay on equilibrium.

inclusion (highly inter-correlated with the differenced independent variables), econometrically acceptable estimates can be obtained by excluding year dummies (see Table 4 and 5).

Consequently, if both the first-difference model and the error correction model allow us to address the potential technical trade-offs in a sufficiently adequate way, which is the appropriate TSCS specification for welfare state development? The answer must be researched in the theoretical assumptions about the welfare state expansion. Because many social programs mature over a long period of time and partisan effects need considerable time to materialize in social spending (Huber and Stephens; Garrett and Mitchell 2001; Kittel and Obinger 2002), a model specification that is able to capture the long-run dynamics must be necessarily adopted. This means that the error correction model is better than the first-difference specification for modeling welfare state development. Its ability to capture both long- and short-term relationships among variables (even when the dependent variable has been differenced) makes the error correction model a good solution to deal with technical and theoretical complications simultaneously. However, if we focus on the government composition impact on social spending exclusively, comparing results of the first-difference with those of the error correction model appears of considerable interest. While the effect of the strength of leftist parties in government during their term in office is statistically significant in the first-difference equations, adding the long-run modelling via an error correction specification involves that both the transitory and permanent effect of the leftist party power turn out to be statistically insignificant. This is substantively remarkable. In fact, although it is reasonable to assume that partisan effect concretises in welfare spending over a long period of time, that impact seems empirically relevant if and only if we model the short-run effects solely.

Bibliography:

Achen, C.F., 2000: Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variable,. Conference paper. Annual Meeting of the Political Methodology Section of the American Political Science Association, UCLA, July 20–22.

Avelino Filho, G., Brown, D.S. and Hunter W.A., 2002, Globalization, Democracy and Social Spending in Latin America, 1980–1987, Paper to be delivered at the Third Meeting of the Associação Brasileira de Ciência Política (ABCP). Niterói.

Baltagi, B.H., 2002, Econometric Analysis of Panel Data, John Wiley and Sons.

Beck N., 2001, Time-Series-Cross-Section Data: What Have We Learned in the Past Few Years?, Annual Review of Political Science, Vol. 4, 271-93.

Beck, N., 1991, Comparing Dynamic Specification: The Case of Presidential Approval, in Political Analysis, Vol. 3, 51-88.

Beck, N. and Katz J.N, 1996, Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models, Political Analysis, Vol. 6, 1-36.

Beck N. and. Katz J.N, 1995, What To Do (and Not To Do) with Time-Series Cross-Section Data, in American Political Journal Review, Vol. 89, 634-47.

Durr, R., 1993, An Essay on Cointegration and Error Correction Models, Political Analysis, Vol. 4, 185-228.

Fox, J. 1991, Regression Diagnostics, Sage.

Garrett, G., 1998, Partisan Politics in the Global Economy, Cambridge, Cambridge University Press.

Garrett, G. and Mitchell, D., 2001, Globalization, Government Spending and Taxation in the OECD, European Journal of Political Research, Vol. 39, 145-77.

Granger, C.W.J. and Engle, R.F., 1987, Cointegration and Error Correction: Representation, Estimation and Testing, Econometrica, Vol. 55, 251-76.

Granger, C.W.J. and Newbold P, 1974, Spurious Regressions in Econometrics, in Journal of Econometrics, Vol. 2, 111-20.

Hicks A., 1994, Introduction to Pooling, in T. Janoski and A. Hicks (edited by), The Comparative Political Economy of the Welfare State, Cambridge University Press.

Hicks, A.M. and Swank, D.H., 1992, Politics, Institutions, and Welfare Spending in Industrialized Democracies, 1960-82, in American Political Science Review, Vol. 86, 658-74.

Hicks, A. and Misra, J., 1993, Political Resources and the Growth of the Welfare in Affluent Capitalist Democracies, 1960-1982, in American Journal of Sociology, Vol. 99(3), 668-710.

Hsiao C., 1986, Analysis of Panel Data, Cambridge University Press.

Huber E., and Stephens J.D., 2001, Development and Crisis of the Welfare State: Parties and Policies in Global Markets, University of Chicago press.

Huber E., Ragin C., and Stephens J.D., 1998, Comparative Welfare States data Set, Northwestern University and University of North Carolina.

Huber E., Ragin C., and Stephens J.D., 1993, Social Democracy, Christian Democracy, Constitutional Structure, and the Welfare State, in American Journal of Sociology, Vol. 99, 711-49.

Iversen, T. and Cusack T., 2000, The Causes of Welfare State Expansion. Deindustrialization or Globalization?, World Politics, Vol. 52, 313–49.

Janoski, T., and Isaac, L.W., 1994, Introduction to time-series analysis, in T. Janoski and A. Hicks (edited by), The Comparative Political Economy of the Welfare State, Cambridge University Press, 31-53.

Judge G.G., Griffiths W.E., Hill R.C., Lutkepohl H. and Lee T-C., 1985, The Theory and Practice of Econometrics, New York, Wiley, 2nd Ed.

Kmenta, J., 1990, Elements of Econometrics. New York: Macmillan; London: Collier Macmillan, 2nd Ed.

Kittel B. and Winner H., 2002, How Reliable is Pooled Analysis in Political Economy? The Globalization-Welfare State Nexus Revisited, MPIfG Discussion Paper 02/3, Cologne.

Kittel B. and Obinger, H., 2002, Political Parties, Institutions, and the Dynamics of Social Expenditure in Times of Austerity, MPIfG Discussion Paper 02/1, Cologne.

OECD, 1998, OECD Statistical Compendium, (on cd-rom), Paris, OECD.

Parks R.W., 1966, Efficient Estimation of a System of Regression Equation When Disturbance are Both Serially and Contemporaneously Correlated, in Journal of the American Statistical Association, Vol. 62, 500-09.

Schmidt M.G., 1997, Determinants of Social Expenditure in Liberal Democracies, in Acta Politica, Vol. 32, 153-73.

Swank D.H., 1992, Structural Power and Capital Investment in the Capitalists Democracies, in American Political Journal Review, Vol. 86, 38-54.

Wooldridge J.M., 2002, Introductory Econometrics: A Modern Approach, South-Western College Publishing, 2nd Ed.

Table 1: Variables

Mnemonic	Description	Source
SSTRAN	social security transfers/GDP (%)	OECD (1998)
LEFTCUM	Cumulative cabinet score for Left parties from 1946 to the year of the observation	Huber et al (1998)
RELCUM	Cumulative cabinet score for religious parties from 1946 to the year of the observation	Huber et al (1998)
RGDPC	GDP per capita	OECD (1998)
DEPRATIO	Population under 15 and over 64 year as percentage of total population	OECD (1998)
TRADE	Export plus import as percentage of GDP	OECD (1998)
UNEM	Unemployment rate	OECD (1998)

Table 2: Determinants of the Levels of Social Security Transfers according to the Huber/Stephens and the Garrett/Mitchell specification strategy

	(1)	(2)	(3)	(4)	(5)
SSTRAN t-1		0.953***	0.994***	0.913***	0.986***
		(0.031)	(0.012)	(0.035)	(0.012)
UNEM	0.463***	-0.001	0.003	0.056**	0.002
	(0.038)	(0.020)	(0.012)	(0.023)	(0.015)
DEPRATIO	-0.030	0.011	0.000	0.072***	0.013
	(0.079)	(0.022)	(0.014)	(0.026)	(0.014)
LEFTCUM	0.305***	0.002	0.009**	0.020**	0.010**
	(0.027)	(0.011)	(0.004)	(0.010)	(0.004)
RELCUM	0.247***	-0.003	-0.002	-0.017	0.002
	(0.045)	(0.016)	(0.005)	(0.017)	(0.005)
TRADE	-0.007	-0.019***	-0.000	-0.005	-0.000
	(0.009)	(0.006)	(0.002)	(0.007)	(0.002)
RGDPC	0.000	-0.000***	-0.000	0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	465	465	465	465	465
R squared	0.433	0.988	0.987	0.983	0.982
F (time effect)		35***	111818***		
F (country effect)		25899***		25**	
Durbin “m” test	0.938***	0.208***	0.199***	0.265***	0.213***
	(0.015)	(0.053)	(0.048)	(0.052)	(0.048)
Mean VIF	1.77	10.88	2.50	10.40	2.26
Rho	0.935				
Modified Wald test chi2 (15)	1236.64***	108.43***	100.86***	63.44***	61.95***
Breusch-Pagan LM test	1209.244***	164.654***	195.705***	375.420***	394.594***

Notes: Standard errors in parentheses.
Significant at 10%; ** significant at 5%; *** significant at 1%.
Coefficients for Fixed Effects not shown.
Dubin “m” test for autocorrelation in residuals (Kmenta 1990).
Mean VIF (Variance inflation factors): indicator of multicollinearity).
Modified Wald test for groupwise heteroskedasticity.
Breusch-Pagan LM test for cross-correlation of residuals.

Table 3: Panel Unit Roots Tests according to the Kittel/Winner specification strategy

Social Security Transfers, Levels				
Levin & Lin	Coefficient	t-value	t	p
Constant	-0.116	-5.721	-1.291	0.098
Constant and trend	-0.232	-7.418	0.668	0.748
Im, Pesaran & Shin	t-bar	cv-10	Ψ	p
Demeaned, no trend	-1.460	-1.810	-0.260	0.397
Demeaned, trend	-1.993	-2.440	1.139	0.127
Not demeaned, no trend	-0.821	-1.810	3.020	0.001
Not demeaned, no trend	-2.090	-2.440	0.405	0.343

Social security transfers, Changes				
Levin & Lin	Coefficient	t-value	t	p
Constant	-0.812	-13.051	-3.856	0.000
Constant and trend	-0.848	-13.589	-4.235	0.000
Im, Pesaran & Shin	t-bar	cv-10	Ψ	p
Demeaned, no trend	-3.248	-1.820	-7.345	0.000
Demeaned, trend	-3.326	-2.450	-5.211	0.000
Not demeaned, no trend	-3.427	-1.820	-8.107	0.000
Not demeaned, no trend	-3.441	-2.450	-5.731	0.000

Notes: Levin & Lin (1993) (LL) tests augmented by 1 lag, H0: nonstationarity. coefficient: Coefficient on lagged levels.t-value: t-value of coef. t*: transformed t-value, ~N(0,1). p: p-value of t*.
Im-Pesaran-Shin (1997) (IPS) tests augmented by 1 lag, H0: nonstationarity. t-bar: mean of country-specific Dickey-Fuller tests.
cv10: 10% critical value of IPS test. Ψ: transformed t-bar statistic, ~N(0,1). p: p-value of Ψ.

Table 4: Determinants of the Changes of Social Security Transfers according to the Kittel/Winner specification strategy

	(1)	(2)	(3)	(4)
Δ SSTRAN t-1	0.027	0.065	0.034	0.060
	(0.058)	(0.058)	(0.055)	(0.055)
Δ UNEM	0.146***	0.143***	0.195***	0.192***
	(0.043)	(0.043)	(0.041)	(0.042)
Δ DEPRATIO	0.035	0.034	0.108	0.094
	(0.081)	(0.081)	(0.091)	(0.089)
Δ LEFTCUM	0.089	0.081	0.129*	0.106**
	(0.075)	(0.054)	(0.074)	(0.053)
Δ RELCUM	-0.013	-0.017	-0.087	-0.043
	(0.173)	(0.100)	(0.179)	(0.098)
Δ TRADE	-0.011	-0.012	-0.013	-0.014
	(0.011)	(0.011)	(0.009)	(0.009)
Δ RGDPC	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	450	450	450	450
R squared	0.552	0.523	0.480	0.450
F (time effect)	1615***	4608***		
F (country effect)	75***		65***	
Durbin “m” test	-0.002	0.077	0.060	0.140*
	(0.098)	(0.099)	(0.081)	(0.081)
Mean VIF	2524828450222	3460381203919	2.25	1.35
Modified Wald test chi2 (15)	236.99***	187.24***	308.72***	172.02***
Breusch-Pagan LM test	27.917*	127.923*	147.046***	146.220***

Notes: Standard errors in parentheses.
Significant at 10%; ** significant at 5%; *** significant at 1%.
Coefficients for Fized Effects not shown.
Burbin “m” test for autocorrelation in residuals (Kmenta 1990).
Mean VIF (Variance inflation factors): indicator of multicollinearity).
Modified Wald test for groupwise heteroskedasticity.
Breusch-Pagan LM test for cross-correlation of residuals.

Table 5: Determinants of Social Security Transfers according to the Iversen/Cusack specification strategy

	(1)	(2)	(3)	(4)
SSTRANt-1	-0.069***	-0.084***	-0.009	-0.014
	(0.027)	(0.025)	(0.011)	(0.010)
ΔUNEM	0.169***	0.238***	0.155***	0.209***
	(0.041)	(0.036)	(0.041)	(0.037)
UNEM t-1	0.004	0.009	-0.027**	-0.036***
	(0.019)	(0.017)	(0.011)	(0.009)
ΔDEPRATIO	0.050	0.062	-0.004	0.027
	(0.092)	(0.097)	(0.085)	(0.090)
DEPRATIO t-1	0.038*	0.063***	-0.003	0.004
	(0.020)	(0.021)	(0.011)	(0.010)
ΔLEFTCUM	0.070	0.114	0.049	0.087
	(0.075)	(0.075)	(0.071)	(0.069)
LEFTCUM t-1	0.007	0.008	0.000	-0.002
	(0.009)	(0.008)	(0.004)	(0.004)
ΔRELCUM	0.014	0.030	0.136	0.098
	(0.185)	(0.189)	(0.182)	(0.180)
RELCUM t-1	-0.010	-0.017	-0.004	-0.001
	(0.014)	(0.014)	(0.006)	(0.006)
ΔTRADE	-0.009	-0.013	-0.006	-0.013
	(0.012)	(0.008)	(0.011)	(0.008)
TRADE t-1	-0.006	0.001	0.002	0.003*
	(0.005)	(0.004)	(0.002)	(0.002)
ΔRGDPC	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
RGDPC t-1	0.000	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	450	450	450	450
R squared	0.581	0.532	0.534	0.480
F (time effect)	74945.4***		1.7e+05***	
F (country effect)	62.22***	75.00***		
Durbin “m” test	0.029	0.031	0.074	0.069
	(0.056)	(0.054)	(0.052)	(0.050)
Mean VIF	2.25e+12	9.65	2.97e+12	2.40
Modified Wald test chi2 (15)	272.97***	305.57***	210.37***	170.06***
Breusch-Pagan LM test	138.322**	136.902**	126.076*	131.693**

Notes: Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1% Coefficients for Fized Effects not shown.
Burbin “m” test for autocorrelation in residuals (Kmenta 1990).
Mean VIF (Variance inflation factors): indicator of multicollinearity.
Modified Wald test for groupwise heteroskedasticity.
Breusch-Pagan LM test for cross-correlation of residuals.