Panel data analysis in comparative politics: Linking method to theory

THOMAS PLÜMPER[1], VERA E. TROEGER[2] & PHILIP MANOW[3]
[1]*University of Konstanz, Germany;* [2]*Max Planck Institute for Research into Economic Systems, Jena, Germany;* [3]*Max Planck Institute for the Study of Societies, Cologne, Germany*

Abstract. Re-analyzing a study of Garrett and Mitchell ('Globalization, government spending and taxation in the OECD', *European Journal of Political Research* 39(2) (2001): 145–177), this article addresses four potential sources of problems in panel data analyses with a lagged dependent variable and period and unit dummies (the de facto Beck-Katz standard). These are: absorption of cross-sectional variance by unit dummies, absorption of time-series variance by the lagged dependent variable and period dummies, mis-specification of the lag structure, and neglect of parameter slope heterogeneity. Based on this discussion, we suggest substantial changes of the estimation approach and the estimated model. Employing our preferred methodological stance, we demonstrate that Garrett and Mitchell's findings are not robust. Instead, we show that partisan politics and socioeconomic factors such as aging and unemployment as expected by theorists have a strong impact on the time-series and cross-sectional variance in government spending.

Introduction

Some years ago Gary King (1990: 11) suggested evaluating the choice of a method in statistical analysis by asking the question: 'What did it do to the data?' He claimed: 'Knowing that one's procedure meets some desirable statistical criteria is comforting but insufficient. We must fully understand (and communicate) just what was done to the data to produce the statistics we report.' In our view, applying the 'King-criterion' is especially warranted in the analysis of what has become one of the standard tools of applied researchers in comparative politics: panel data. Yet the choice of panel data analysis has often not been theoretically justified, but rather bypassed with a general reference to the 'standard' that has emerged in the field over the last years. Although panel analysis leaves room for a wide variety of model specifications, one method suggested by Nathaniel Beck and Jonathan Katz, here called the 'de facto Beck-Katz standard', has become the accepted econometric technique in comparative political economy (Beck & Katz 1995, 1996; Beck 2001).

This technique suggests running an OLS regression on panel data, adding the lagged dependent variable plus unit and period dummies to the battery of right-hand side variables, and calculating 'panel corrected standard errors'.[1]

A standard never, however, emerges without some dangers attached. Researchers become overconfident that they have chosen the 'correct' method or they feel forced to employ 'standard techniques' in order to get their research published, even if the use of such techniques may not be fully justified. In this article, we want to highlight some of the dangers caused by a mismatch between theory and methodology due to the unwarranted use of an 'econometric standard'. We do so by re-analyzing the recently published article by Geoffrey Garrett and Deborah Mitchell, who wrote: 'We are confident about the veracity of [our] empirical claims given the cautious methodological stance we have adopted' (Garrett & Mitchell 2001: 175; see Appendix A for a replication of their estimates). We will show that their confidence was premature since their use of the Beck-Katz method cannot qualify as a 'cautious methodological stance'.

In particular we will discuss:

- whether the use of unit dummies can lead to the absorption of most of the theoretical interesting *cross-sectional variance* in the data;
- whether the inclusion of a lagged dependent variable and period dummies absorbs most of the theoretically interesting *time-series variance* in the data;
- how much results in first difference models depend on the chosen lag-structure; and
- whether the assumption of time-constant coefficients is supported by the data.

Before beginning the discussion, we will provide the reader with a brief introduction to panel data analysis.

The de facto Beck-Katz standard

Analyzing data and testing theories in comparative political economy is hampered by the notorious 'few cases, many variables' problem. At best, about 100 countries in the world publish reasonably complete and reliable data on their political, social and economic system in a regular fashion. For many research questions of high theoretical interest, the number of cases is even smaller. An obvious solution to the small-N problem has been the analysis of pooled time-series cross-section data (hereafter, just 'pooled data'). In fact, 'pooling' has become the most prominent and promising remedy for the small-N problem

in political science. It is so widely used in comparative political economy that 'it has become difficult to defend *not*' using it (Kittel & Winner 2003: 5; emphasis added).

'Pooling' seems to offer at least two advantages compared to either a pure cross section of units or a pure time series. First, the number of observations increases and so do the degrees of freedom. This makes it possible to estimate more fully specified models. Second, pooling makes it possible to control for exogenous shocks common to all countries (by controlling for time effects) and to reduce the omitted variable bias (by controlling for unit effects).

However, the analysis of pooled time-series data is more problematic than a pure cross-section analysis because observations are usually not independent. There are four potential violations of OLS standard assumptions in panel data:

- errors tend to be autocorrelated (serial correlation of errors) – that is, they are not independent from one time period to the other;
- errors tend to be heteroscedastic – that is, they tend to have different variances across units, for instance units with higher values may have a higher error variance (panel heteroscedasticity);
- errors tend to be correlated across units due to common exogenous shocks (contemporaneous correlation of errors); and
- errors may be non-spherical in both the serial and the cross-sectional dimension (i.e., autocorrelated and heteroscedastic at the same time).

To deal with panel heteroscedasticity, Beck and Katz (1995) proposed a procedure that adopts robust standard errors in an OLS estimate. However, the OLS estimator is known to be biased if errors are autocorrelated. Therefore, the procedure suggested by Beck and Katz necessarily requires that one eliminates serial correlation of errors. They recommend adding the lagged dependent variable to the right-hand side of the equation, which in almost all cases suffices to get rid of autocorrelation. Beck and Katz (1996: 4) worked with fixed effects models to control for the possibility of non-spherical errors in the time and cross-sectional dimension (a fixed effects model includes dummies for N-1 units and T-1 periods).

A typical estimation equation then looks as follows:

$$y_{i,t} = \alpha + \beta_0 y_{i,t-1} + \sum_{k=1}^{K} \beta_k x_{k,i,t} + \varepsilon_{i,t} \tag{1}$$

where β_0 is the autoregression coefficient and where

$$\varepsilon_{i,t} = \upsilon_i + \vartheta_t + v_{i,t} \tag{2}$$

with $\alpha + \upsilon_i$ being the unit specific intercept and ϑ_t controls for period effects and $v_{i,t}$ measures the non-systematic error.

In combination, the inclusion of the lagged dependent variable, the time and unit dummies and the panel corrected standard errors deal with all four potential problems. Hence, from an econometric perspective there is nothing wrong with the de facto Beck-Katz standard. However, problems may arise with respect to the theory-method nexus.[2] Four of these problems will be addressed in the following sections: the unjustified absorption of cross-sectional variance, the unjustified absorption of time-series variance, the mis-specification of the lag structure and the unjustified assumption of time-invariance of causal relations.

Country dummies, time invariance and cross-sectional variance

We begin our discussion of the de facto Beck-Katz standard by taking up the concerns that have most often been raised by applied researchers and econometricians: the pros and cons of controlling for unit fixed effects.[3] Country fixed effects can be estimated by

$$y_{i,t} = \alpha + \upsilon_i + \sum_{k=1}^{K} \beta_k x_{k,i,t} + \varepsilon_{i,t}, \tag{3}$$

where $\alpha + \upsilon_i$ is the unit specific intercept, with υ_i estimated by N-1 country dummies and where α is the intercept of the -1st country.

Two concerns about the consequences of a fixed effects specification have been brought up in comparative politics. First, unit dummies have been accused of eliminating 'too much' cross-sectional variance (Huber & Stephens 2001); and second, the inclusion of unit dummies makes it impossible to estimate the effect of time invariant exogenous variables (Wooldridge 2002) and severely biases the estimate of partly time invariant variables (Beck 2001). We consider these points in turn.

Often the inclusion of country fixed effects is recommended because the coefficients of unit dummies are interpreted as measures of *unobserved* time invariant variables. This interpretation finds support by econometricians. For instance, Peter Kennedy (1998: 227) states: 'The dummy variable coefficients reflect ignorance – they are inserted merely for the purpose of measuring shifts in the regression line arising from unknown variables.' Yet, whether to include or exclude country dummies has stimulated a lively debate among political scientists. Garrett and Mitchell have taken a firm stance. They believe that country dummies have to be included to account for the 'underlying histori-

cal fabric of a country that affects the dependent variable and that is not captured by any of the time and country-varying regressors' (Garrett & Mitchell 2001: 163). At the other end of the spectrum Kittel and Obinger (2002: 21) object that unit fixed effects throw 'out the baby with the bath water, because one of the main interests of political scientists in this kind of quantitative analyses is whether institutional variables capture cross-sectional variation to an extent which makes the inclusion of country dummies unnecessary'. If one examines what the inclusion of unit dummies does to the data, one finds neither the hopes of Garrett and Mitchell nor the fears of Kittel and Obinger fully confirmed.

Country dummies control away the deviation of the variables' mean of one unit from the variables' mean of the base unit. Thus, unit dummies *completely absorb* differences in the *level* of independent variables across units. This effect is dubbed the 'within-transformation'. As a consequence, the country dummy variable can be considered as a vector of the effects of time invariant variables plus some parts of the mean of time varying variables. To see why this happens and to understand what dummies do to the data, consult the estimated variance-covariance matrix, which is used to estimate coefficients in a fixed-effects model (the matrix is usually called 'Q-matrix'):

$$Q_T y_i = Q_T x_{k,i}\beta_1 + Q_T(e_T z_i \beta_2) + Q_T \varepsilon_i \qquad (4)$$

where z_i is a vector of time-invariant variables and

$$Q_T y_i = |y_{i,1} - \bar{y}_i, \ldots, y_{i,T} - \bar{y}_i| \qquad (5a)$$

$$Q_T x_{k,i} = |x_{k,i,1} - \bar{x}_{k,i}, \ldots, x_{k,i,T} - \bar{x}_{k,i}| \qquad (5b)$$

$$Q_T \varepsilon_i = |\varepsilon_{i,1} - \bar{\varepsilon}_i, \ldots, \varepsilon_{i,T} - \bar{\varepsilon}_i| \qquad (5c)$$

where Q denotes a matrix and

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{i,t} \qquad (6)$$

The estimation equation is thus given by

$$y_{i,t} - \bar{y}_i = \beta_k (x_{k,i,t} - \bar{x}_{k,i}) + (\varepsilon_{it} - \bar{\varepsilon}_i) \qquad (7)$$

First, the equations show that the potential effect of time invariant variables is completely ignored (time invariant z variables do not appear in the estimation equation 7). In other words, the estimated coefficient of the time invariant z variables is zero, which may not be the true effect. Second, the unit dummies estimate the effect of the time invariant variables together with some effects of the differences in levels of the time varying variables *plus* the mean

effect of omitted variables. Hence, Garrett and Mitchell's interpretation of dummy coefficients as capturing the historical fabric of a country is highly problematic. Third, the unit fixed effects measure also the pre-t_1 effect of the right-hand side variables as soon as the dependent variable reveals variance that already existed at t_1 (i.e., in the starting period of the panel). The existence of significant unit dummies therefore cannot necessarily be taken as an indicator for the existence of omitted variables. This means that Kittel and Obinger's claim that a model is only fully specified if country fixed effects are no longer different from zero ignores the fact that usually the period under observation does not begin with the first day of variance in the dependent variable.

Abstention from running a fixed effects model comes with the risk of explaining variance in the dependent variable that existed *prior* to the period under observation with the variance in the mean of the independent variable *in* the period under observation. This procedure may be acceptable if the largest effect on the dependent variable comes from independent variables that are historically given constants such as geography. However, if the independent variables vary over time, then the exclusion of country dummies may easily lead to biased estimates and wrong inferences because the variance in the dependent variable that came into being before t_0 is explained by the independent variable's average variance in levels between t_0 and T. Obviously, if one wants to explain the variance of government spending that was generated before t_0, one preferably should collect the appropriate historical data. If this is not possible, it is certainly better not to make inferences.

We cannot simply assume that the causes of the observed increase in government spending between 1960 and 1993 must be equal to the causes before 1960. Nor can we easily assume that, on average, the distribution of values in the independent variables between 1960 and 1993 is a good proxy for the distribution of values in the same variables before 1960. If we used the observable variance in growth rates between 1960 and 1993 to explain the variance in government spending that already existed in 1960, we would implicitly assume that countries having comparably high growth rates between 1960 and 1993 also had the highest growth before 1960.

Since this assumption cannot be taken for granted, the elimination of cross-sectional differences is not a mistake and not as problematic as it may seem at a first glance. Quite to the contrary, since the fixed effects model exclusively uses the variance in the unit dynamics of the independent variables to estimate the dynamics of the dependent variable, it is the estimation procedure of choice for a wide variety of research questions.

Having said this, there is one crucial exception where the inclusion of country dummies actually complicates the interpretation of the estimation. If

either a *level effect* of at least one variable or a time invariant variable of theoretical interest exist, the inclusion of country dummies is problematic because it suppresses the level effects.[4] Let us give an example for this problem. As we have seen, the within estimator only estimates the effect of deviations from the group mean $\bar{x}_i$ in the independent variables on the dependent variable. It follows that an increase in LEFT from 0 to 5 per cent and an increase in LEFT from 45 to 50 per cent have the same effects on government spending.[5] The fact that the second country has a much higher participation of left parties in government does not (by implicit assumption) turn up in the estimated parameter of LEFT if we control for unit fixed effects. However, from a theoretical point of view it is plausible, that the *level* of left cabinet seats exerts an influence on *changes* in government spending. Another way to state this effect is by saying that it has absolutely no influence on the estimation of the Garrett-Mitchell model if we, for instance, replace the value 0 for left cabinet share in the case of the United States by the value 100. It leaves the estimated coefficient of LEFT unchanged.[6] This seemingly crucial manipulation of the data only reduces the coefficient of the United States dummy by 100 times the coefficient of LEFT. Do we really believe that United States government spending today would be the same if the country had been continuously governed by a left government? At least, this is what Garrett and Mitchell implicitly assume. The same (i.e., nothing) would happen to the estimated coefficients of the substantive variables if we generously added 5 percentage points of per capita growth to all European countries. Even linear transformations in opposed directions have absolutely no impact – thus, if the United States had always been governed by left parties and the participation of left parties in the Swedish government had always been 25 percentage points lower, we would still obtain the same estimated coefficients (with the exception of the country dummies).

In short, all discriminatory and non-discriminatory linear transformations only influence the dummy's coefficients, which can easily be demonstrated mathematically:

$$x_{k,i,t} - \bar{x}_{k,i} = (x_{k,i,t} + a) - (\bar{x}_{k,i} + a) \tag{8}$$

A linear transformation of all values of x for one unit affects the observed values and the mean value equally. Obviously, the difference between the observed value of a variable and the unit average remains constant. Since only the difference between the observed value of x and the mean value of x is used for the estimation, the estimated coefficient remains the same.

From a theoretical perspective, this exclusive reliance on *changes* in levels is problematic. If we assume that it should make a difference for government spending and for the changes in government spending whether a country has

an average left cabinet share of 0 per cent (as in the case of the United States) or 73.5 per cent (as in the case of Sweden), then the effect of this difference is hidden in the vector of influences that make up the unit fixed effects coefficient.[7] In another re-analyses of the Garrett-Mitchell paper, Bernhard Kittel and Hannes Winner (2003: 13) conclude that the fixed effects specification 'cannot clearly confirm the presence of partisan effects'. Our conclusion is different. Left parties in governments (may) have a persistent level effect on changes in government spending. In the short-run, however, changes in the participation of left parties in government do not necessarily matter much. Thus, there can be no general rule to either include or exclude country dummies. The specification of an estimation model depends completely on the theory the researcher wants to test. If one believes in persistent level effects, one has to find a way around the inclusion of unit dummies even if this means that estimates are biased by omitted variables.[8]

To sum up, if the theory says something about level effects on levels or on changes, a fixed effects specification is not the model at hand. If a theory predicts level effects, one should not include unit dummies. In these cases, allowing for a mild bias resulting from omitted variables is less harmful than running a fixed effects specification. Alternatively, researchers can choose the two-step approach suggested by Hausman and Taylor (1981, see also Wooldridge 2002: 325ff.). The results of both procedures are biased (the Hausman-Taylor procedure should be less biased), but they at least allow the estimation of level effects and thus have the lowest cost-benefit ratio. Even if researchers are not interested in level effects or the influence of time invariant variables, fixed effects models generate inefficient[9] and possibly biased estimates if T is small and if most of the variance in the dependent variables did already existed in t_1. If an increase in T is impossible or if t_1 is determined by data availability, one cannot obtain unbiased parameter estimates by analyzing panel data. However, due to omitted variable bias, one can also not calculate unbiased estimates by suppressing country dummies or running a pure cross section.

Lagged dependent variables and period dummies, or: What determines the trend in spending levels?

One of the main advantages of pooled data analyses is that pooling allows for explicit modeling of the *dynamics* inherent in the panel (Beck & Katz 1995: 5; Kittel & Winner 2003: 15).[10] Time-series cross-sectional (TSCS) data analyses promise to give answers to questions many researchers are really interested in when they run cross-sectional regressions. However, as we

demonstrate in this section, these hopes may not be fulfilled if researchers adhere to the de facto Beck-Katz standard.

We will show that, under some conditions, the inclusion of period dummies and a lagged dependent variable may cause serious bias. The problem is that common features of panel models that follow the de facto Beck-Katz standard may absorb large parts of the trend without actually explaining it if the dependent variable exhibits a general time trend. Under this circumstance, estimates are biased if at least one variable has persistent effects, while the others do not. If this persistence is not carefully modeled, the coefficient of the lagged dependent variable is biased upwards, while the coefficients of the other independent variables are likely to be biased downward (because the lagged dependent variable wrongly presupposes an identical persistent effect of all independent variables). This argument is not novel. Some econometricians and applied researchers are opposed to the inclusion of a lagged dependent variable to control for serial correlation, at least if the level of a variable is not influenced by its level in the previous year (Achen 2000; Huber & Stephens 2001: 59).

There are good reasons, however, to believe exactly that. Garrett and Mitchell, for instance, justify the inclusion of the lagged dependent variable by arguing that its coefficient is a measure for the path dependency of the dependent variable. At least in analyzing the dynamics of government spending, it seems only natural to assume persistency:

> Budgets of welfare institutions are made with reference to the budget of the previous year and the largest shares of social spending (health care and old age pensions) tend to increase incrementally. Hence there are substantive reasons for persistence and we should expect that the levels of social expenditure do depend on past levels. (Kittel & Obinger 2003: 24)

Yet the interpretation of the lagged dependent variable as measure of time persistency is misleading. Since the lagged dependent variable can be written as a function of the lagged right-hand side variables, the lagged dependent variable's coefficient measures the weighted average of the right-hand side variables. Therefore, the lagged dependent variable not only evaluates the level of persistency. It also models the dynamics of the independent variables (Cochrane & Orcutt 1949). In this regard, the lagged dependent variable implicitly assumes that the dynamics of all independent variables are identical. Needless to say, this assumption is not very convincing and almost certainly wrong. Thus, the inclusion of the lagged dependent variable 'seems more

an afterthought than a reasoned model specification decision firmly grounded in theory' (Wawro 2002: 47).

Rather than being included for theoretical reasons, the lagged dependent variable's inclusion seems to be the result of methodological imperatives, given that it addresses the problem of serial correlation in the errors (Beck & Katz 1995; Maddala 1998: 63). We will show in the following paragraphs that modeling the dynamics in a panel by the lagged dependent variable may bias the estimates because any trend in the dependent variable will be absorbed by theoretically uninteresting variables such as the lagged dependent variable, the period dummies and the intercept. The argument is developed in two steps, first showing how much of the trend is absorbed by the non-substantively interesting variables, and second, arguing why the coefficient of the lagged dependent variable is likely to be biased upwards in the presence of persistent effects and a trend-ridden dependent variable.

To begin with, the impact of the non-substantial variables on a time trend can be calculated by a polynomial following

$$y_t(\alpha, v) = \alpha \sum_{n=0}^{T} \beta_0{}^n + \sum_{n=0}^{T} \vartheta_t \beta_0{}^{(t-n)}, \quad n = [0, T] \tag{9}$$

where β_0 is the estimated coefficient of the lagged dependent variable. Equation 9 illustrates how the intercept, the lagged dependent variable and the period dummies commonly capture much of the time-series variance in panel data. Most noteworthy, if researchers exclude one of these elements from the model, large parts of its conditional effect are absorbed by the remaining elements.

This exercise illustrates why we cannot simply interpret the coefficient of the lagged dependent variable as a measure of time persistency. In the Garrett-Mitchell model, the lagged dependent variable's coefficient is 0.914. Implicitly, Garrett and Mitchell suggest that on average, 91.4 per cent of government spending is persistent, but this interpretation pays no attention to the dynamic effects. At t_n the aggregated effect of the lagged dependent variable is already $\sum_{n=1}^{t_n} \beta_0^{t-n}$. In other words, at t_5 the aggregated effect of the lagged dependent variable is $\beta_0 + \beta_0^2 + \beta_0^3 + \beta_0^4 + \beta_0^5 \approx 3.849$. At t_{22}, the lagged dependent variable alone already accounts for an increase in government spending of about 9.15 per cent of the average country's GDP. In addition to this effect, we have to take the dynamic effects of the constant plus the unit fixed effects plus the period fixed effects into consideration. Both effects also sum up if a lagged dependent variable is included.

Note that the inclusion of a lagged dependent variable implies that researchers assume divergence in spending levels across countries. Since the

absolute increase in the dependent variable explained by the non-substantive variables depends on the initial level in the dependent variables, these variables predict a larger increase in government spending the higher the initial level. This implicit assumption is counterintuitive and not theoretically grounded.

Since it may be difficult to understand this trend-absorbing effect of non-substantive control variables, we visualize the absorbed trend in Figure 1. In the period under observation, the average share of government spending has increased from slightly below 30 to almost 55 per cent. This increase is represented by the grey-shaded area in Figure 1. The five lines represent common model specifications that are included to control for unobserved effects in panel data. The first model specification, represented by the downward triangle, includes only period dummies. Since period dummies can easily follow the trend, this estimation procedure removes the trend by about 80 per cent. The second variant, represented by the square, adds the lagged dependent variable to the period dummies. The parameter of the period dummies is by and large reduced by the cumulated effect of the lagged dependent variable. Garrett and Mitchell use this procedure. Due to the fact that the effect of previous period dummies is sustained by the lagged dependent variable (especially when its coefficient approaches 1.0), this estimation procedure is particularly difficult
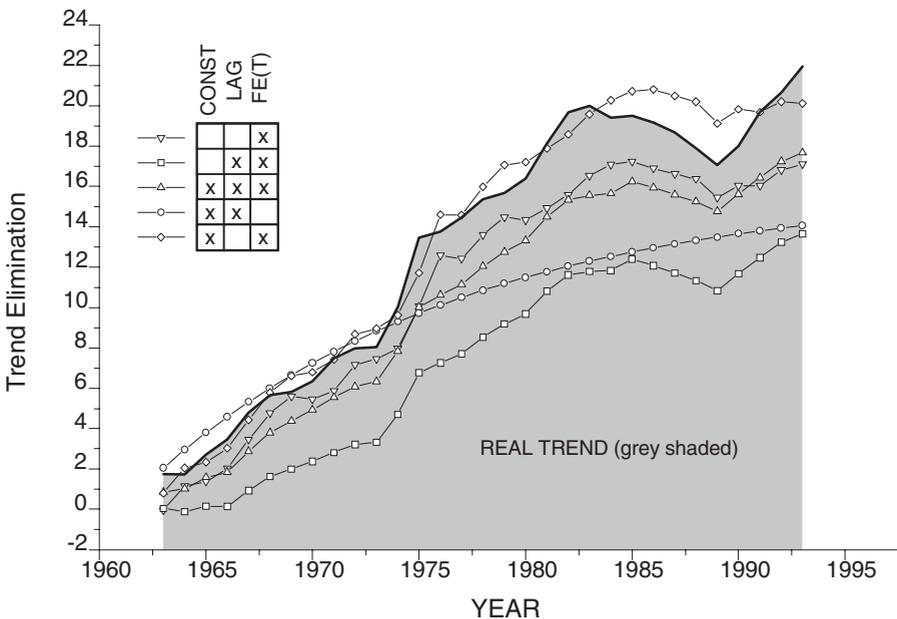

Figure 1. Trend 'elimination' in panel data.

to interpret. With regard to the trend in government spending, the model specification has the advantage of absorbing less variance than all other variants. The third model augments the second by inclusion of the constant. It approaches the trend line following a function similar to the period dummies procedure. The fourth variant excludes period fixed effects from the estimation equation, but includes the lagged dependent variable and an intercept for obvious reasons. The exclusion of period fixed effects brings about a monotonically increasing, well-behaved function. Finally, the fifth variant represented by the diamonds, excludes the lagged dependent variable. This version eliminates the trend almost completely. If one had to choose among these variants in analyzing government spending in OECD countries, the version employed by Garrett and Mitchell leaves the comparatively largest share of the variance for the exogenous variables.

The five variants are highly correlated with each other and with the observed values (all correlation coefficients are larger than 0.90). What does this exercise tell us? First of all, it is important to see that the estimates of the lagged dependent variable, the constant and the period dummies depend to a very high degree on the model specification. Perhaps most importantly for the theoretical discussion of exogenous variables, the inclusion of the lagged dependent variable and/or time dummies leaves very little variance for the explanatory variables.

As a consequence of the method they have chosen, Garrett and Mitchell propel a large part of the dynamics in government spending into the lagged dependent variable and the time dummies. The corollary of this procedure becomes obvious by comparing the Garrett-Mitchell model to a theoretical empty model that includes nothing but the lagged dependent variable plus the country and time dummies. From an econometric point of view, the various goodness-of-fit indicators point out that the empty model is not worse. From a theoretical point of view, however, the empty model has no substance since it 'explains' the total variance exclusively by unobserved factors such as unknown country and time effects.

If we consider this issue from the assumption that the model specification chosen by Garrett and Mitchell is unbiased, the case is clear. The variables identified by Garrett and Mitchell are not those that bring about the steep rise in government spending. If, on the other hand, we do not take the unbiasedness of the de facto Beck-Katz model for granted,[11] the result may look different. Shall we believe that about 90 per cent of the growth of the welfare state is exogenously determined and follows a change in tastes for social security and/or changes in other unobserved variables such as union strength or the increase in the power of bureaucrats? Shall we believe that the variations in the exogenous variables of Garrett and Mitchell's research design explain

less than 10 per cent of the changes in government spending? Can we believe that the increase in unemployment did not contribute significantly to the increase in government spending?[12]

Our answer to these questions is that the coefficient of the lagged dependent variable is indeed biased upwards, if at least one of the independent variables exerts a persistent effect on the dependent variable, which is ignored by researchers. For instance, if the true model is

$$y_{i,t} = \alpha + \beta_0 y_{i,t-1} + \beta_1 x_{i,t} + \beta_2 x_{i,t-1} + \varepsilon_{i,t} \qquad (10)$$

where $\beta_2 > 0$, but researchers estimate

$$y_{i,t} = \alpha + \beta_0 y_{i,t-1} + \beta_1 x_{i,t} + \varepsilon_{i,t} \qquad (11)$$

$\hat{\beta}_0$ is biased upward, because the omitted effect of β_2 influences the estimation of β_0.[13] To add some empirical evidence to the claim that the lagged dependent variable's coefficient is biased upwards, we report the results from a simple cross-sectional OLS regression of the growth of public spending from 1961 to 1993 on the averages in the exogenous variables selected by Garrett and Mitchell.

Table 1 raises some doubts as to whether the size of the lagged dependent variable's parameter in the Garrett-Mitchell model is appropriate. Models 1.1 and 1.2 illustrate that the variables included in the Garrett-Mitchell model explain more than 50 per cent of the difference in a pure cross section. It is

Table 1. Cross-sectional analysis of the trend (dependent variable: spending in 1994–spending in 1961)

	Model 1.1	Model 1.2
Y-intercept	5.2950 (1.1830)***	5.4860 (1.2700)***
Average unemployment	0.0778 (0.0273)**	0.0678 (0.0278)**
Average growth	−0.0120 (0.0762)	−0.0613 (0.0801)
Average dependency ratio	−0.1498 (0.0349)***	−0.1532 (0.0377)***
Average left	0.0085 (0.0024)***	
Years with left portfolio > 50%		0.0192 (0.0061)**
Average Christian Democrats	−0.0079 (0.0026)**	−0.0075 (0.0027)**
Average trade	0.0056 (0.0031)	0.0087 (0.0034)**
Average low-wage imports	−0.0056 (0.0099)	−0.0012 (0.0109)
Average FDI	−0.0713 (0.0660)	−0.1114 (0.0701)
N	17	17
Adj. R^2	0.604	0.549
F-stat	4.05**	3.43*

Note: ***$p < 0.01$; **$p < 0.05$; *$p < 0.10$.

true that due the cross-sectional specification of the model we are not able to control for exogenous shocks and our model is prone to omitted variable bias. This is certainly problematic and we do not take the estimated coefficients at face value; but we take the results from a pure cross section as a hint that the model specification chosen by Garrett and Mitchell substantially underestimates the importance of the increase in the unemployment rate and of the policies implemented by left governments for changes in government spending.

It is beyond the task of this article to give a definitive answer with respect to the determinants of public spending. Instead we ask: Do different understandings of the causes for the postwar increase in government spending matter at all? In order to calculate the impact of the autocorrelation coefficient on the coefficients of the theoretically interesting exogenous variables, we propose fixing the coefficient of the lagged dependent variable at different values. Specifically, we use the following formula to convert the vector of the dependent variable

$$\tilde{y}_t = y_t - y_{t-1} \cdot \tilde{\beta}_0 \tag{12}$$

where $\tilde{\beta}_0$ is the (arbitrarily fixed) autocorrelation coefficient and $\tilde{y}_0$ is the share of government spending that remains unexplained given our enforced autocorrelation coefficient.

The reader may feel tempted to ask if it is justified to set a prior of the autocorrelation coefficient. The answer in our case is yes and no at the same time. Methodologically, the answer is definitely positive (cf. Beck & Katz 2001: 15). From a theoretical perspective, however, the answer remains controversial simply because there is no generally accepted prior for the parameter of an autoregressive process in government spending.[14] A solution to this problem may be, as Beck and Katz (2001: 30) eventually came to conclude

> that we have to allow ourselves more judgment. If we as political scientists really believe that the units are relatively homogeneous, we can impose that in our priors. But if we move from being either empirical Bayesians or Bayesians with very gentle priors, we of course run the risk of others doubting our work because they disagree with our priors. The solution here, as suggested by Bartels, is that we provide a variety of results, allowing our priors to vary. It may be that the Bayesian estimates vary a lot with choice of prior, in which case we have to worry about which prior we believe.

Table 2 reports the results for three arbitrarily chosen $\tilde{\beta}_0$ values (models 2.1 to 2.3), while model 2.4 reports the results for the $\tilde{\beta}_0$ estimated by Garrett

Table 2. The effect of priors of the LDVs coefficient

	Model 2.1 FE(C) $\tilde{\beta}_0 = 0.000$	Model 2.2 FE(C) $\tilde{\beta}_0 = 0.300$	Model 2.3 FE(C) $\tilde{\beta}_0 = 0.700$	Model 2.4 FE(C) $\tilde{\beta}_0 = 0.914$	Model 2.5 FE(C) $\tilde{\beta}_0 = 0.000$	Model 2.6 FE(C) $\tilde{\beta}_0 = 0.300$	Model 2.7 FE(C) $\tilde{\beta}_0 = 0.700$
Unemployment	1.4860 (0.0846)***	1.0130 (0.0610)***	0.3810 (0.0369)***	−0.0434 (0.0336)	1.0410 (0.0978)***	0.7730 (0.0759)***	0.2940 (0.0442)***
GDP per capita growth	−0.8250 (0.1044)***	−0.7110 (0.0720)***	−0.5600 (0.0355)***	−0.4790 (0.0299)***	−0.3390 (0.0309)***	−0.3800 (0.0361)***	−0.4520 (0.0315)***
Dependency ratio	0.4963 (0.0288)***	0.3694 (0.0204)***	0.2004 (0.0119)***	0.1101 (0.0110)***	0.5566 (0.0444)***	0.4109 (0.0322)***	0.2159 (0.0176)***
Left cabinet portfolios	−0.0094 (0.0059)	−0.0064 (0.0042)	−0.0025 (0.0023)	−0.0003 (0.0019)	0.0051 (0.0045)	0.0042 (0.0038)	0.0025 (0.0026)
Christian Democrat portfolios	−0.0400 (0.0090)***	−0.0286 (0.0065)***	−0.0134 (0.0040)***	−0.0052 (0.0036)	−0.0088 (0.0081)	−0.0075 (0.0070)	−0.0043 (0.0046)
Trade openness	0.2060 (0.0283)***	0.1380 (0.0204)***	0.0473 (0.0122)***	−0.0011 (0.0111)	0.0308 (0.0245)	0.0265 (0.0216)	0.0103 (0.0149)
Low-wage imports	−0.0271 (0.0393)	−0.0224 (0.0283)	−0.0161 (0.0155)	−0.0013 (0.0120)	−0.0156 (0.0300)	0.0028 (0.0245)	0.0203 (0.0140)
Foreign direct investment	0.6059 (0.1853)***	0.3981 (0.1281)***	0.1210 (0.0622)*	−0.0269 (0.0499)	0.1150 (0.0979)	0.1095 (0.0843)	0.0583 (0.0583)
N	529	529	529	529	529	529	529
R²	0.991	0.991	0.988	0.939	0.822	0.805	0.770
Wald χ^2	5006619.43	4667513.45	629304.69	58831.88	279377.87	385942.12	348954.87
Prob. > χ^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PCSE	yes	yes	yes	yes	yes	yes	yes
Time dummies	no	no	no	no	no	no	no
Country dummies	yes	yes	yes	yes	yes	yes	yes
Error correction	no	no	no	no	AR1	AR1	AR1
Rho					0.707	0.620	0.313

Notes: *** p < 0.01; * p < 0.10.

and Mitchell to allow an easy comparison of the adjusted R^2 and the levels of significance. Note that similar to the conversion of country fixed effects in the previous section, the conversion of the vector of the dependent variable does not influence the coefficients, but the standard errors.[15] As mentioned above, the differences in the standard errors result from the increase in the degrees of freedom, here caused by the exclusion of the lagged dependent variable.[16]

However, the serial correlation of the residuals increases with a lower weight on the autoregressive process and OLS estimates become inefficient. We therefore combine panel-corrected standard errors with a Prais-Winsten transformation (AR1)[17] to eliminate serial correlation of errors and report the results in models 2.5 to 2.7. Note that these models – even model 2.5, which does not include a lagged dependent variable – neither show spherical distribution of errors nor do they fail a test of autocorrelation of errors. Thus, they satisfy the minimum requirements for an OLS estimate. This is not the case with models 2.1 to 2.3, which are biased due to serial correlation of errors.

Table 2 reveals the dependence of the exogenous variables' estimated coefficients on the coefficient of the lagged dependent variable. The higher the share of the a-theoretically absorbed time-series variance, the worse the model and the less significant the variables of theoretical interest. It immediately becomes obvious that – as the coefficient of the lagged dependent moves toward 1.0 – most coefficients of the independent variables approach zero. Thus, the prior belief about the exogeneity of the trend has an impact on the interpretation of results. If, for instance, we assume the trend to be completely exogenous, neither partisanship nor international economic integration influences government spending. If, however, we believe that the trend is at least 30 per cent a consequence of the variables included in the Garrett-Mitchell model, we must conclude that Christian Democratic governments tend to reduce government spending while trade openness and long-term capital flows increase the public share of the economy. Thus, a proper estimation of the trend's degree of exogeneity is crucial. Of course, one can believe that everything works fine in the Beck-Katz model and that the results that Garrett and Mitchell have observed reveal the true influences on government spending. If this were true, the increase in government spending must have been brought about by factors not properly included by the variables of the Garrett-Mitchell model. The model was thus a poor approximation of the dynamics of government spending. In our view, the elimination of serial correlation by inclusion of the lagged residuals gives more appropriate coefficients than the inclusion of a lagged dependent variable. Most importantly, AR1 error models are consistent, while the inclusion of a lagged dependent variable renders estimates

necessarily inconsistent if random or fixed effects are included (Wooldridge 2002: 270).

To sum up this section's argument, the lagged dependent variable in possible combination with period dummies and an intercept captures large parts of the trend in the dependent variable. Though researchers may interpret this result as a hint that the independent variables do not explain the trend, we have argued that the coefficient of the lagged dependent variable is biased upwards. This becomes plausible if one compares the results from a pure AR1 error model, which leaves much more variance for the substantive exogenous variables. In principle, AR1 error models tend to absorb less time-series dynamics and may therefore be the method of choice for applied researchers seeking to explain not only cross-sectional variance and cross-sectional differences in changes, but also average changes in levels. A good compromise seems to be the use of a theoretical prior to fix the coefficient of the lagged dependent variable. Again our discussion demonstrates that the least harmful specification of the estimation model clearly depends on the theory of the researcher.

A unique political reaction function? The time-lag problem

Up to this point, we have discussed the rationale for controlling country and time fixed effects as well as for the inclusion of a lagged dependent variable. While these problems were explicit assumptions of the de facto Beck-Katz standard, there also exists the implicit assumption of *uniform lags*. This problem is related to the issue of time invariant variables. If changes in the independent variables are rare and of little persistence, a good estimate critically depends on the selection of a correct lag structure. This is well known from time-series analyses.[18]

In panel data analyses, the additional difficulty emerges that the correct lag structure can be unit and/or time dependent. Especially in estimating the effect of institutional changes on a dependent variable, lags may seriously differ from country to country. Thus, whenever a research agenda addresses the impact of institutional factors on policies or on social and economic outcomes, researchers should take into account the possibility that lags are not identical across units:

> The assumption of uniform leads and lags lies at the heart of any pooling endeavor except if the leads and lags are estimated separately for each country and variable – not to mention the possibility of variation between particular subperiods within countries – which implies that we tend to

> end up, for example, with a set of single-country time-series regressions. (Kittel & Obinger 2002: 26)

Since the estimation of unit dependent lags is difficult and time-consuming, most researchers either do not lag the independent variables or arbitrarily choose uniform lags – most often one-year lags.

We will use a simple AR1 error model to illustrate the lag structure's impact on the coefficient and the level of significance of one variable: LEFT. We have selected this variable because jumps in the value of LEFT are quite frequent, but persistent. Every time there is a cabinet reshuffle, the share of left cabinet seats may change. Most importantly, however, political scientists should not expect uniform reaction functions of governments in countries with different institutional settings (Tsebelis 2002; Döring 1995).[19] If institutions matter, they not only influence socioeconomic variables, but also delay the reaction of governments in countries with less government autonomy.

There is no generally accepted indicator for the determination of the length of the lag, but there are several candidates: the t-statistic, the R^2, the AIC (Akaike Information Criterion) and the BIC (Bayesian Information Criterion). Since we are less interested in finding the 'true' lag structure than in demonstrating that the choice of a lag structure significantly influences the parameter estimates, we used an un-weighted composite index statistic as a selection criterion. Though this procedure is an obvious simplification, it is nevertheless much more appropriate than the simplifying assumption of a unique lag structure for all observations.

As a result, we get eleven countries where a change in government has an immediate effect on government spending. In four countries (Australia, Austria, Germany and Ireland), we found a one-year lag; in two countries (Italy and Denmark), a two-year lag; and in another two countries (Finland and Netherlands), a three-year lag performed best. Though the optimization of lags is certainly time consuming, it is absolutely essential in first difference models.

Rejecting a hypothesis because the estimated coefficient turned out to be insignificant while in fact the researcher has wrongly assumed uniform lags would mean blaming the theory for the failures of the methodology. In the long run, however, political scientists should be able to formulate expectations on the length of the lag using the inputs of constitutional and institutional political economy. Table 3 illustrates the fact that the optimization of lags renders estimates significant without any other change in the model specification.

Compare the AR1 error model with uniform lags (model 3.1) to the model in which the unlagged left cabinet portfolio variable is replaced by the optimized lag version of LEFT (model 3.2). A comparison of the results shows that the estimated effect of the optimized lag is more than 40 times larger

Table 3. Arbitrarily chosen versus optimized lags: A comparison

	Model 3.1: uniform lags	Model 3.2: optimized lag
Unemployment	1.2146 (0.8830)***	1.2168 (0.0860)***
GDP per capita growth	−0.2604 (0.0403)***	−0.2608 (0.0401)***
Dependency ratio	−0.7384 (0.1267)***	−0.6655 (0.1229)***
LEFT (no lags)	0.0002 (0.0041)	
LEFT (optimized lags)		0.0083 (0.0038)**
Christian Democrat portfolios	−0.0418 (0.0093)***	−0.0338 (0.0088)***
Trade openness	0.0507 (0.0290)*	0.0515 (0.0285)*
Low-wage imports	−0.1488 (0.0436)***	−0.1567 (0.0420)***
Foreign direct investment	−0.0910 (0.0951)	−0.1217 (0.8604)
N	529	524
R^2	0.940	0.944
Wald χ^2	13035.32	12310.68
Prob. $> \chi^2$	0.0000	0.0000
PCSE	yes	yes
Time dummies	no	no
Country dummies	yes	yes

Notes: ***$p < 0.01$; **$p < 0.05$; *$p < 0.10$.

(0.0083 instead of 0.0002) than the unlagged effect of LEFT, while the standard deviation remains approximately constant. This supports the belief that a rejection of the partisan effect hypothesis based on the insignificance in the conventional first difference model is not well founded. In the next section, we present another criticism leading to the same result: partisan effects can be detected in the data, if researchers take the theory as seriously as the method.

Is 'left policy' stable over time? The implicit assumption of constant parameter values

It is a heroic assumption that parameter values remain constant over time (Kennedy 1998: 107).[20] Indeed, Maddala (1998: 67) sees the neglect of what he called 'structural changes' as one of the major shortcomings of panel data analysis. Structural breaks can take different forms, of which changes in slopes and changes in error variances rank most prominently. In this section, we deal exclusively with changes in slopes. The reason is straightforward. No empirical analysis of the determinants of public spending that we are aware of

discusses and controls for changing parameter values, despite the numerous theoretical studies claiming that left policies have changed over time and governments dominated by left parties no longer implement a Keynesian macroeconomic policy. To control for this widely accepted hypothesis, researchers have to allow for parameter heterogeneity; if they do not, they are not taking the contending theories of the welfare state seriously enough.

To keep the empirical model as easy as possible, we demonstrate the necessity to set the coefficients of the partisan variables free. In comparing the PCSE model to the AR1 error model, we can demonstrate not only that the impact of, especially, left parties on government spending does not remain constant over time, but we can also provide evidence that the autocorrelation in errors significantly decreases after allowing parameter values to vary over time. Specifically, the estimated coefficients of models with and without included lagged residuals become much more similar if one allows for parameter heterogeneity.

In principle, there are two procedures that allow parameter values to differ. The first variant identifies periods P (one can of course also use T), adds period dummies to the dataset and creates interaction effects between all period dummies and the variable under suspicion of exerting an unstable effect over time. In this case, the pure variable (here: LEFT) is excluded from the regressors. In the second variant, the suspicious variable remains included. Therefore, one can only include P-1 periods in the model. The second variant is preferable because it allows us to see if the differences of included periods relative to the base period are significant. For simplicity, we declare the first period to be the base period. Note, however, that making an outlying period the base renders all differences significant, while making the average period the base leaves other periods more often than not insignificant.

Table 4 nicely illustrates the importance of testing for parameter heterogeneity. Two results are noteworthy. First, the left party influence on the size of the public sector has increased on average from the early 1960s to the late 1980s in our sample. Second, Christian Democratic parties, on average, increased government spending approximately as much as left governments did. However, they reversed their positive impact on spending significantly earlier. This result can be more easily seen in Figure 2 (coefficients are taken from model 4.3).

To sum up, we cannot ignore the possibility of a structural change in slopes if theories predict time varying parameter values. It therefore does not come as a surprise that we were able to estimate coefficients for left and Christian Democrat cabinet shares that were significantly different to those of Garrett and Mitchell. Allowing for time variant influences improves the model, halves the serial correlation of errors and confirms theories that are well established.

Table 4. Testing for parameter heterogeneity

	Model 4.1 FE(C) $\beta_0 = 0.7$	Model 4.2 FE(C) $\beta_0 = 0.7$	Model 4.3 FE(C) $\beta_0 = 0.7$	Model 4.4 FE(C) $\beta_0 = 0.7$	Model 4.5 FE(C) $\beta_0 = 0.7$	Model 4.6 FE(C) $\beta_0 = 0.7$
Unemployment	0.3737 (0.0361)***	0.3897 (0.0370)***	0.3770 (0.0366)***	0.3918 (0.0514)***	0.4041 (0.0524)***	0.3861 (0.0510)***
GDP per capita growth	−0.5349 (0.0332)***	−0.5559 (0.0354)***	−0.5284 (0.0330)***	−0.4334 (0.0327)***	−0.4385 (0.0338)***	−0.4308 (0.0330)***
Dependency ratio	0.2004 (0.0114)***	0.1987 (0.0129)***	0.2036 (0.0126)***	0.2020 (0.0171)***	0.2037 (0.0188)***	0.2066 (0.0183)***
Left cabinet portfolios	−0.0356 (0.0042)***	−0.0031 (0.0023)	−0.0355 (0.0041)***	−0.0271 (0.0054)***	−0.0013 (0.0029)	−0.0276 (0.0053)***
Left 1966–1970	0.0204 (0.0046)***		0.0198 (0.0045)***	0.0127 (0.0053)**		0.0130 (0.0051)**
Left 1971–1975	0.0343 (0.0044)***		0.0334 (0.0042)***	0.0275 (0.0056)***		0.0274 (0.0055)***
Left 1976–1980	0.0417 (0.0047)***		0.0406 (0.0047)***	0.0296 (0.0062)***		0.0303 (0.0061)***
Left 1981–1985	0.0451 (0.0050)***		0.0459 (0.0051)***	0.0372 (0.0066)***		0.0389 (0.0066)***
Left 1986–1990	0.0445 (0.0051)***		0.0444 (0.0051)***	0.0354 (0.0069)***		0.0368 (0.0068)***
Left 1991–1994	0.0408 (0.0069)***		0.0384 (0.0066)***	0.0330 (0.0080)***		0.0311 (0.0078)***
Christian Democrat portfolios (CDEM)	−0.0078 (0.0040)*	−0.0261 (0.0064)***	−0.0183 (0.0063)***	−0.0072 (0.0056)	−0.0205 (0.0090)**	−0.0161 (0.0086)*
CDEM 1966–1970		0.0079 (0.0059)	0.0024 (0.0058)		0.0051 (0.0092)	0.0003 (0.0085)
CDEM 1971–1975		0.0142 (0.0081)*	0.0099 (0.0077)		0.0175 (0.0131)	0.0131 (0.0119)
CDEM 1976–1980		0.0304 (0.0071)***	0.0261 (0.0068)***		0.0217 (0.0111)**	0.0201 (0.0103)**
CDEM 1981–1985		0.0108 (0.0090)	0.0120 (0.0088)		0.0138 (0.0119)	0.0146 (0.0114)
CDEM 1986–1990		0.0113 (0.0078)	0.0119 (0.0078)		0.0129 (0.0105)	0.0128 (0.0100)
CDEM 1991–1994		0.0157 (0.0079)**	0.0198 (0.0079)**		0.0236 (0.0111)**	0.0256 (0.0103)**
Trade openness	0.0305 (0.0120)**	0.0397 (0.0139)***	0.0210 (0.0142)	0.0288 (0.0163)*	0.0347 (0.0179)*	0.0204 (0.0180)
Low-wage imports	0.0306 (0.0156)**	0.0280 (0.0154)*	0.0393 (0.0157)**	0.0100 (0.0213)	0.0078 (0.0205)	0.0186 (0.0205)
Foreign direct investment	0.0439 (0.0566)	0.1469 (0.0601)**	0.0567 (0.0583)	0.0077 (0.0709)	0.0579 (0.0703)	0.0061 (0.0707)
N	529	529	529	529	529	529
R^2	0.989	0.988	0.990	0.698	0.676	0.709
Wald χ^2	1413549.32	487709.39	1192927.23	308660.36	304701.82	340991.67
Prob. > χ^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PCSE	yes	yes	yes	yes	yes	yes
Time dummies	no	no	no	no	no	no
Country dummies	yes	yes	yes	yes	yes	yes
Error correction	no	no	no	AR1	AR1	AR1
Rho				0.408	0.415	0.394

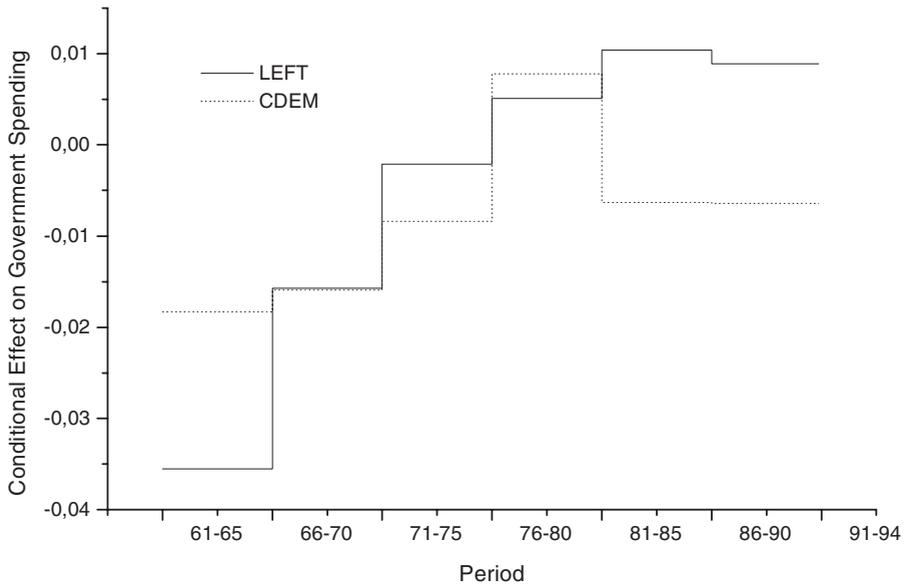Notes: *** p < 0.01; ** p < 0.05; * p < 0.10.

Figure 2. The conditional effects of LEFT and CDEM in government spending.

Ignoring these effects in the presence of well-founded theories leads to unsound inferences.

Conclusion

The preceding discussion led to the following results. First, and partly contrary to what many comparative political economists believe, unit fixed effects turn out to be problematic if variables are time invariant or if the theory at test predicts *level effects*. The first case has been widely acknowledged in the literature. It is problematic because while a simple fixed effects specification does not work, an *unbiased* and *efficient* alternative estimation procedure does not exist. In the absence of a first best solution, the Hausman-Taylor procedure provides the second best solution. The second case has probably led to more wrong inferences simply because it is less known. If the theory predicts an influence of the level of an exogenous variable on changes in the endogenous variable, unit fixed effects *must* not be included. It is most important to note that unit dummies already bias estimates if at least one exogenous variable exerts a persistent level effect on the dependent variable. However, most researchers add unit dummies on an ad hoc basis because they know of the existence of a standard procedure, but not what this procedure does to the

data. As a consequence, regression analyses often do not test the theory researchers claim to test.

Second, the inclusion of a lagged dependent variable and/or period dummies tends not only to absorb large parts of the trend in the dependent variable, but likely biases estimates. To the extent researchers are interested in explaining this trend, panel data analysis based on the methodology suggested by Beck and Katz (1995, 1996), especially if based on the inclusion of the lagged dependent variable, may not be appropriate. We suggest that researchers use the Prais-Winsten transformation rather than the lagged dependent variable to eliminate serial correlation of errors whenever the dependent variable is trend-ridden and the researcher beliefs that the explanatory variables can explain the trend.

Third, simply assuming a uniform lag structure may cause biased estimates and wrong inference. We not only found that the assumed length of the lag matters, it is at least equally important to test for the option that the political process is not equally efficient in different countries and, if this test is confirmed, to allow for unit varying lags.[21] As yet, a standard operating procedure for estimating the unit-specific lag length in a panel does not exist. The procedure we have used in this article is analytically sound, but very time consuming.

Fourth, if the time dimension in panel analyses exceeds a rather limited number of time periods, it becomes extremely important to think about and test for structural changes in slopes and error variance. This is especially true for the impact of political institutions and actors. Consider the impact of parties on macroeconomic outcomes. There is no *a priori* reason to believe that coefficients should be stable over time. To the contrary, it is very likely that coefficients are time dependent. Since these effects can easily be controlled for, there are no reasons not to do so.

Which general conclusion follows from the preceding analysis? We think it has become clear that the results derived from panel data analysis critically depend on a host of crucial methodological decisions and theoretical assumptions which the use of a 'standard' like the de facto Beck-Katz method cannot avoid, but only hide. This article has sought to reveal these often hidden assumptions and question their theoretical justification in our concrete example: the investigation into the determinants of public spending. By doing no more than increasing the method-theory fit, our findings support a number of hypotheses that have frequently been formulated in the comparative political economy of the welfare state and are *not* in accordance with the results derived by Garrett and Mitchell. First and foremost, we found that partisan effects matter. However, party preferences' influence on government spending is not stable over time. Second, our results say that unemployment and the aging of the society tend to put an upward pressure on government budgets,

while growth reduces the government share of the economy. Third, international economic openness does not seem to have a similarly important influence on government spending. We would like to emphasize that these findings stand in contrast to Garrett's and Mitchell's results gained on the basis of the same data, but with a different methodological approach.

Acknowledgments

Appendix A. Replication of Garrett and Mitchell's estimates

	Model 1.1: Garrett/ Mitchell model	Model 1.2: Results reported by Garrett & Mitchell
Government spending$_{t-1}$	0.9140 (0.0190)***	0.911
Unemployment	0.0214 (0.0388)	0.033
GDP per capita growth	−0.3650 (0.0319)***	−0.362
Dependency ratio	0.0837 (0.0453)*	0.097
Left Cabinet portfolios	−0.0023 (0.0020)	−0.003
Christian Democrat portfolios	−0.0042 (0.0037)	−0.005
Trade openness	−0.0258 (0.0110)**	−0.030
Low-wage imports	0.0266 (0.0165)	0.023
Foreign direct investment	−0.0072 (0.0565)	0.006
N	529	529
R^2	0.999	0.999
Wald χ^2	33652.78	
Prob > χ^2	0.000	
PCSE	yes	yes
Time dummies	yes	yes
Country dummies	yes	yes

Notes: ***p < 0.01; **p < 0.05; *p < 0.10.

Notes

1. It should, however, be stressed that Beck and Katz did not necessarily advocate the de facto standard that developed after publication of their articles on panel analysis.
2. However, we know that including unit-dummies and a lagged dependent variable at the same time renders the estimation inconsistent.
3. In order to save space, we do not discuss a random effects specification. This can be easily be justified because: if T becomes large, differences between random and fixed effects estimation becomes small; an F-test more often reveals the superiority of running a fixed effects model; and random effects are very likely to be inconsistent.
4. Estimates are even more biased if the level of an independent variable influences the variance in the dependent variable.
5. Problematic here is *not* the assumption of linearity. If the researcher doubts that the effect of absolute changes of one independent variable on the dependent variable is linear, they should not abstain from running a fixed effects model, but rather allow for non-linearities in the functional form by adding the square of the independent variable to the list of regressors.
6. If we drop the unit dummies, the manipulation of the US LEFT values has the expected consequence: it reduces the value of the coefficient.
7. If you do not believe that differences in levels can have an impact on a time varying variable, take a trip up a mountain and measure the temperature. If you repeat this, you will notice that levels do not only have an influence on your dependent variable, but also on the variance in the dependent variable and thus on their changes. Ignoring this fact causes biased estimates.
8. However, this is not the only potential solution. A possible alternative is to run a fixed effects model and then regress the estimated country dummy coefficients on a set of regressors including, of course, those variables that may exert level effects (Plümper & Troeger 2004).
9. Formally, a fixed effect model is inefficient if the level variance of the independent variables exceeds the level variance of the dependent variable.
10. Even if we accept that the Beck-Katz methodology offers a significant improvement over the Parks-Kmenta model, this assertion is not fulfilled. The coefficient of the lagged dependent variable coerces all other variables into an identical autoregressive process, which is implausible for theoretical reasons.
11. See Kittel and Winner (2003: 18) for a more detailed discussion of bias induced by inserting the lagged dependent variable.
12. This position has not remained uncontested. For instance, Evelyn Huber and John Stephens take an opposite point of view. Trying to save as much variance as possible for their independent variables, they lag the dependent variable by five years and do not use period dummies. To justify their procedure, they argue: 'The unstandardized variable coefficent for the [five-year] lagged dependent variable varied from 0.20 to 0.47. While this is considerably lower than the 0.92 to 0.97 with the one-year time lag, the lagged dependent variable was still the most important variable. . . . Based on our case studies, this would appear too high an estimate for the policy legacies effect and thus is probably inflated' (Huber & Stephens 2001: 67). Their skepticism about the appropriateness of the lagged dependent variable coefficient in part results from they way they interpret this coefficient. In their view, 'a 1% higher level of social expenditure will result in 0.2% increase in expenditure five years later' (Huber & Stephens 2001: 67). This

would mean that a 1 per cent higher government share in the Garrett-Mitchell model increases government spending by 0.914 per cent in the following year. Fortunately for the welfare state (and for the results derived by Garrett and Mitchell), Huber and Stephens interpretation is not correct. As we have demonstrated above, the cumulated effect of the lagged dependent variable and the period dummies *exceed 1.0* in the Garrett-Mitchell model, and in all other panel specifications we have estimated so far.

13. The same holds true for persistent effects of omitted variables. For instance, Garrett and Mitchell do not control for the effect of per capita income on government spending. Including this variable reduces the estimated value of β_0.

14. We have experimented with allowing for time and unit varying β_0 values, thereby modelling the idea that political degrees of freedom vary over time and from country to country. Though it can be shown that β_0 varies across units and over time, we do not report these results here. Note, however, that the United States, Switzerland, Australia, New Zealand, Great Britain, Ireland and Japan have relatively low autocorrelation coefficients (high political degrees of freedom); Canada, France, Germany and Austria have medium autocorrelation coefficients (medium political degrees of freedom); while the other countries have significantly higher autocorrelation coefficients. Contrary to Beck and Katz's claim that it is not very sensitive to assume unit specific ρ, these findings make theoretical sense and confirm widely shared beliefs about 'families of nations' (Castles 1994).

15. The differences between the coefficients estimated by model 1 (the Garrett-Mitchell model) and model 12 thus are the sole consequence of the exclusion of the time dummies. If we had included the time dummies in model 12, we would have obtained identical coefficients. There is, however, a good reason not to include the time dummies: if we had, the time trend would have been absorbed by them and we would not gain additional variance.

16. Since the lagged dependent variable correlates with the dependent variable, the inclusion of a lagged dependent variable causes an upward bias in the standard errors. This bias is largest for the variables that perform best in regard to the t-statistics. Thus, the inclusion of a lagged dependent variable may cause under-confidence in results. We therefore recommend checking the severity of the bias by converting the dependent variable according to Equation 7 and dropping the lagged dependent variable and (in case of a trend in the dependent variable) the time dummies from the model.

17. Our estimation procedure models the dynamics as an AR1 process within the error term: $y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}$, where $\varepsilon_{it} = \rho\varepsilon_{it-1} + u_{it}$, u_{it} is assumed to be white noise. Beck (2001: 280) states that 'one solution that should work well is the TSCS variant of the single equation error correction model'.

18. Kittel and Winner (2003: 22) have suggested differencing the dependent variable and the independent variables as a remedy for the non-stationarity (i.e., $\beta_0 \geq 0.9$). Estimating a first difference model, however, makes the choice of the correct lag structure even more important because only the short-run effects of the exogenous variables are estimated. For instance, Kittel and Winner implicitly assume that changes in trade openness in one year cause increases in government spending in the very same year – and have no effect in the following year.

19. Our data-mining approach to the lag length lends support to Herbert Döring's (1995: 225) analysis of agenda setting power and control of the plenary timetable, as well as to George Tsebelis' (2002) veto-player approach. Both, Döring's 'authority to determine plenary agenda' variable and Keefer's 'checks1' variable (Beck et al. 2000) show a sig-

nificant and robust effect on the lag length we have estimated. Results and data can be obtained from the authors upon request.

20. Since we analyze a fairly homogenous sample, we ignore potential estimation problems that may arise from heterogeneous unit samples (see Bartels 1996, Western 1998).

21. Again, there is no free lunch. The danger of fixed effects models results from reversed causality.

References

Achen, C.H. (2000). *Why lagged dependent variables can suppress the explanatory power of other independent variables* (Political Methodology Working Paper). Available online at: http://polmeth.wustl.edu/papers/00/achen00.pdf.

Bartels, L.M. (1996). Pooling disparate observations. *American Journal of Political Science* 40(3): 905–942.

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.

Beck, N. & Katz, J. (1995). What to do (and not to do) with time series cross-section data. *American Political Science Review* 89(3): 634–647.

Beck, N. & Katz, J. (1996). Nuisance versus substance: Specifying and estimating time-series cross-section models. *Political Analysis* 6(1): 1–36.

Beck, N. & Katz, J. (2001). Random Coefficients Models for Time-series Cross-section Data. Unpublished manuscript, University of California at San Diego.

Beck, T. et al. (2000). *New tools and new tests in comparative political economy: The database of political institutions*. Washington: World Bank.

Castles, F.G. (1993). *Families of nations: Patterns of public policy in Western democracies*. Aldershot: Dartmouth.

Cochrane, D. & Orcutt, G. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* 44(1): 32–61.

Döring, H. (1995). Time as a scarce resource: Government control of the agenda. In H. Döring (ed.), *Parliaments and majority rule in Western Europe*. New York: St. Martin's Press, pp. 223–246.

Garrett, G. & Mitchell, D. (2001). Globalization, government spending and taxation in the OECD. *European Journal of Political Research* 39(2): 145–177.

Hausman, J.A. & Taylor, W.E. (1981). Panel data and unobserved individual effects. *Econometrica* 49(6): 319–339.

Huber, E. & Stephens, J.D. (2001). *Political choices in global markets: Development and crisis of advanced welfare states*. Chicago, IL: Chicago University Press.

Kennedy, P. (1998). *A guide to econometrics*, 4th edn. Cambridge, MA: MIT Press.

King, G. (1990). On political methodology, *Political Analysis* 2: 1–29.

Kittel, B. & Obinger, H. (2002). *Political parties, institutions and welfare state dynamics in times of austerity* (Discussion Paper 1/2002). Cologne: Max-Planck Institut für Gesellschaftsforschung.

Kittel, B. & Winner, H. (2003). *How reliable is pooled analysis in political economy: The globalization-welfare state nexus revisited* (Discussion Paper 3/2003). Cologne: Max-Planck Institut für Gesellschaftsforschung.

Maddala, G.S. (1998). Recent developments in dynamic econometric modeling: A personal viewpoint. *Political Analysis* 7(1): 59–87.

Plümper, T. & Troeger V. (2004). *The estimation of time-invariant variables in panel analyses with unit effects* (Polmeth and SSRN Working Paper). Available online at: http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=374864.

Tsebelis, G. (2002). *Veto players: How political institutions work*. New York: Sage.

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

Western, B. (1998). Causal heterogeneity in comparative research: A Bayesian hierarchical modeling approach. *American Journal of Political Science* 42(4): 1233–1259.

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

Address for correspondence: Thomas Plümper, University of Konstanz, Box D 86; D-78457 Konstanz, Germany

Tel.: +49-7531-88 2608/3081; Fax: +49-7531-88 2774; E-mail: Thomas.Pluemper@uni-konstanz.de