Is there a selection bias in roll call votes? Evidence from the European Parliament

Is there a Strategic Selection Bias in Roll Call Votes in the European Parliament?

Revised. 22 July 2014

Simon Hix
London School of Economics and Political Science

Abdul Noury
New York University

Gerard Roland
University of California, Berkeley

Keywords: roll call votes, European Parliament, party discipline, natural experiment, difference in difference estimation

JEL codes: P16, P48, D72, D78

Abstract: Carrubba, Murrah, Clough, Montgomery and Shambaugh (2006) argue that there is a strategic selection bias in roll call votes in the European Parliament, and provide evidence of some significant differences between roll call votes and other votes in the European Parliament, such as the over-representation of some policy areas, non-legislative votes, and requests by some political groups in roll call votes. In this research note, we examine the magnitude and significance of selection bias in roll call votes due to strategic motives. Prior to 2009, all roll call votes in the European Parliament had to be requested by political groups. Since 2009, in contrast, a roll call vote has been required in all final legislative votes. Using a difference-in-difference design, we exploit this change in the European Parliament's rules of procedure, to compare the effects of strategically-requested and exogenously required roll call votes on political group voting cohesion. Using data from the first 18 months of the Sixth (2004-09) and Seventh (2009-14) Parliaments, we do not find any significant differences in the levels of voting cohesion of the main political groups in requested and non-requested roll call votes. These results suggest that selection biases in roll call votes due to strategic choices are at best negligible. We also compare our results with a similar analysis by Mühlböck and Yordanova (2014).

In an article in this journal Carrubba, Gabel, Murrah, Clough, Montgomery and Shambaugh argued that roll call votes (RCVs) in the European Parliament are a biased sample of the total population of votes in that legislature.[1] Carrubba and his team collected information on all votes, including non RCVs, in the European Parliament between July 1999 and June 2000, and discovered several aggregate differences between RCVs and the total population of votes in that period. First, RCVs in the European Parliament are not proportional to the activity of parliamentary committees, since a few committees (Environment, Public Health and Consumer Policy; Citizen's Freedoms and Rights, Justice and Home Affairs; Constitutional Affairs; and Economic and Monetary Affairs) account for slightly over two-thirds of all RCVs but less than a third of all votes. Second, some political groups (especially the Greens) request RCVs more than others, meaning that there is no proportionality between party size and the proportions of RCVs requested.[2] Third, RCVs over-sample non-legislative 'resolutions' relative to legislative bills, as the former constituted over three-quarters of RCVs but counted for less than half of all the votes in the period in question. This evidence hence suggests that where RCVs can be requested by political actors, they are unlikely to be a random sample of all votes. This raises a serious concern for scholars of legislative behaviour, as analysing RCV data may produce biased inferences about such things as the voting cohesion of parties, the types of coalitions that form in legislatures, or which actors and groups are on the winning and losing side most often. Empirically, then, the following question needs to be answered: given that there are good theoretical arguments for why RCVs should be biased for strategic motives, and that there is

[1] Carrubba, Clifford, Matthew Gabel, Lacey Murrah, Ryan Clough, Elizabeth Montgomery and Rebecca Shambaugh, 'Off the Record: Unrecorded Legislative Votes, Selection Bias and Roll-Call Vote Analysis,' *British Journal of Political Science*, 36 (2006), 691-704.

[2] The European People's Party (EPP) was over-represented as a requester of RCVs on final votes whereas the Greens were over-represented as requesters of RCVs on amendment votes.

empirical evidence showing that RCVs are not a random sample of all votes, how biased are roll call votes in practice due to these strategic motives?

At face value, this question seems impossible to answer, since RCVs are by definition the only votes that are observed and recorded.[3] Nevertheless, in this note we aim to contribute a partial answer by exploiting an institutional change in the rules of procedure in the European Parliament between the Sixth (2004-2009) and Seventh (2009-2014) sessions of the legislature – which we refer to as EP6 and EP7 for convenience. Specifically, whereas in EP6 all RCVs had to be requested by one or more of the political groups or 40 Members of the European Parliament (MEPs) in the chamber, from the start of EP7 RCVs became mandatory for all final legislative votes, under Rule 166 of the new Rules of Procedure of the chamber.

To the extent that this change in the rules was unexpected in EP6, and assuming that votes and MEPs can be made comparable across the two periods, it allows us to use a difference-in-differences design to compare the difference between final legislative votes and other votes (amendments and non-legislative votes) in EP6 and EP7, using various controls, including committee topic and identity of the party group sponsoring the bill.[4] By looking at the difference between requested RCVs and non-requested RCVs we are able to filter out any strategic selection effects which are the main alleged cause of biases. We can also take into account differences between conditions of voting in EP6 and EP7 as well as differences between final legislative votes and other types of votes. This is, hence, a useful method for making unbiased estimates in cases where sample selection might potentially be a serious issue.

---

[3] The only exception we are aware of is Hug, who was able to use records from RCV and non-RCV electronic votes in Switzerland to recover individual voting behaviour in votes. He found that party cohesion was on average higher for RCVs than for unrecorded votes. This method cannot be replicated for legislatures where electronic data for non-RCVs is not collected, which at the moment seems to be the case for most legislatures across the world. See Simon Hug, 'Selection Effects in Roll Call Votes,' *British Journal of Political Science*, 40 (2009), 225-35.

[4] Difference-in-differences estimation has become a standard tool in economics and political science. For an exposition see Josh Angrist and Joern-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton, NJ: Princeton University Press, 2009).

A potential objection to our approach might be that the institutional change between EP6 and EP7 is not a natural experiment, because only final legislative votes have become mandatory. One could also argue that conditions in EP7 are not necessarily the same as in the EP6, since the individual MEPs in the two sessions were not identical. However, the difference-in-differences method is relevant because it allows us, under certain conditions, to filter out effects that are specific to final legislative votes relative to other votes; both within EP7 and between EP7 and EP6. We explain, below, how we do this.

Our analysis is closely related to a number of recent studies examining RCV selection effects in the European Parliament.[5] Mühlböck and Yordanova look for the existence of selection bias.[6] They address the question of how voting cohesion might differ between recorded and non-recorded votes as a result of party signalling and/or disciplining efforts. Their analysis indicates that the relative cohesion of political groups on final legislative votes on average increased after the new rules of procedure were introduced. They conclude that relying on RCVs can lead to underestimating rather than overestimating voting cohesion of the political groups in the European Parliament.

Our main finding is that there is no significant difference in political group voting cohesion in the European Parliament when RCVs are mandatory relative to when they are requested by a political group. This result is robust to comparing final legislative votes to only legislative votes and to other RCVs, as well as to the inclusion of various control variables. While there are good theoretical arguments for possible selection biases of RCVs, and requested

[5] Cf. Bjørn Høyland, 'Procedural and Party Effects in European Parliament Roll Call Votes,' *European Union Politics*, 11 (2010), 597-613.
[6] Monika Mühlböck and Nikoleta Yordanova, 'Party Group Cohesion in the European Parliament: Tracing the Bias in Roll-Call Votes', *European Political Science Review* (2014) forthcoming..

RCVs are clearly not a random sample of all votes, our findings suggest that the effect of this selection on aggregate voting behaviour in the European Parliament are negligible.

Recently Mühlböck and Yordanova, M-Y hereafter, published a paper on selection effects in roll call votes using the EP6 and EP7 data. They show that there is a selection effect but contrary to what one would expect, they report a positive effect. There are a number of similarities between our paper and M-Y's. M-Y used very similar data, the same type of analysis, i.e difference-in-differences, to examine the same questions: are there any selection effects in roll call votes? They too rule out any negative selection effect whereby political groups become less cohesive when they are forced to vote publicly. Our analysis differs from theirs in five distinct ways. First, the data sample used is slightly different as we use 60 more final legislative votes than them. As we show below, the difference in sample size plays no role as similar results are obtained using the same method on both samples. Second, they assigned weights to the data. This does not lead to any difference in results either. Third, they used data only for the four main European party groups. Fourth, they used a non linear method, fractional logistic regression. Fifth, they used pooled data whereas we used average cohesion scores over political groups. We also used cohesion scores of each political group separately. These last three differences jointly lead to a positive cohesion effect. If one drops any of these differences compared to our analysis, the result ceases to be significant. Our analysis also explores more controls and robustness checks. We control for the identity of the political group requesting a RCV as well as for the topic voted on, which are both important controls to take into account, given the Carrubba et al. findings mentioned above.[7] Moreover, we also control for composition effects between EP6 and EP7, since there was turnover of individual MEPs between EP6 and

[7] Carrubba et al., 'Off the Record'.

EP7. To do this, we compare the voting behaviour of individual MEPs who were present in both EP6 and EP7 in requested RCVs and non-requested RCVs.

Our main result has important implications for further research on the European Parliament because the absence of a robustly significant difference in cohesion between mandatory and non-mandatory RCVs implies that the systematic use of RCVs in empirical research on legislative behaviour in the European Parliament is useful for policy research.[8]

Having said that, one effect our approach does not preclude is how the transparency implied by RCVs affects the voting behaviour of representatives. That is, our study cannot reveal the effect of transparent voting because we do not have data on the counterfactual: of how individual legislators vote when their voting behaviour is not recorded.

The rest of the paper is organized as follows. Section 1 explains how we use the difference-in-differences approach. Section 2 presents the empirical results, and section 3 concludes.

1. The difference-in-differences approach

The difference-in-differences approach was developed in the context of experiments. Although we are not here in the presence of a natural experiment, we can still use this method if we assume that the votes and voters across EP6 and EP7 and comparable. To make sure that voters (legislators) are identical across time, in one of robustness checks we only keep the MEPs who served in both periods. To render the samples of votes more homogenous, and thus comparable, we include a large number of control variables.

[8] See, for example, Simon Hix, Abdul Noury and Gérard Roland, 'Power to the Parties: Cohesion and Competition in the European Parliament, 1979-2001,' *British Journal of Political Science*, 35 (2005), 209-234; Simon Hix, Abdul Noury and Gérard Roland, 'Dimensions of Politics in the European Parliament,' *American Journal of Political Science*, 50 (2006), 494-511; and Simon Hix, Abdul Noury and Gérard Roland, *Democratic Politics in the European Parliament* (Cambridge: Cambridge University Press, 2007).

Call *T* the "treatment" variable, which takes a value of 1 for the "treatment" group and a value of 0 for the "control" group. In our case, the "treatment" group is composed of final legislative votes (in EP6 and EP7). Call *t* the time variable, which takes a value of 0 before the change in rule,i.e in EP6 and a value of 1 after the rule change, in EP7. Call Y_i the outcome variable, in this case cohesion of a party in a RCV. Call $\overline{Y}_0^T, \overline{Y}_1^T, \overline{Y}_0^C, \overline{Y}_1^C$, respectively, the pre- and post-"treatment" average outcome for the "treated" group and the "control" group: which in our case are the final legislative votes in EP6 and EP7 and non-final legislative votes in EP6 and EP7. The difference-in-differences estimate is then $E\left[\overline{Y}_1^T\right]-E\left[\overline{Y}_1^C\right]-\left(E\left[\overline{Y}_0^T\right]-E\left[\overline{Y}_0^C\right]\right)$, where E[*x*] is the expected value of a random variable *x*. This estimate is obtained either by computing sample means for each category or by performing the following regression:

$$Y_i = \alpha + \beta T_i + \gamma t_i + \delta T_i t_i + \varepsilon_i$$

where ε_i is assumed to be i.i.d with $E(\varepsilon_i)=0$; $cov(\varepsilon_i,T_i)$, $cov(\varepsilon_i,t_i)$, $cov(\varepsilon_i,T_i\ t_i)$ all assumed to be zero. Coefficient δ is hence the difference-in-differences estimate. Indeed,

$$E\left[\overline{Y}_1^T\right]-E\left[\overline{Y}_1^C\right]-\left(E\left[\overline{Y}_0^T\right]-E\left[\overline{Y}_0^C\right]\right)=(\alpha+\beta+\gamma+\delta)-(\alpha+\gamma)-\left[(\alpha+\beta)-\alpha\right]=\delta.$$

Note that the difference-in-differences estimate can be understood here as equivalent to the difference between, on one hand, final legislative votes in EP7 and EP6 and other votes in these two parliaments, or to the difference between final legislative votes and other votes in EP7 and EP6. If we expect political group cohesion in RCVs to be biased upwards due to strategic motives, then δ should be negative since these strategic motives have been eliminated in the treatment group.

Comparing only the final legislative votes in both EP6 and EP7 would not be the correct way of estimating the "treatment" effect, as this would include specific effects due to the

differences between EP6 and EP7 and due to the passage of time. Comparing final legislative votes only in EP7 with other votes in EP7 would also not be correct because it would include specific effects due to final legislative votes. The difference-in-differences method estimates correctly the effect of the change in voting rule while filtering out effects due to the differences between the two parliaments as well as differences between final legislative votes and other votes.

To compare like-for-like we focus on the first 1,350 votes from each parliament (EP6 and EP7). Among these votes some are legislative votes, some are non-legislative resolutions and some are votes on budgetary issues. We use all the votes to explore the variations in voting behaviour over time. In our baseline regressions, we use only legislative votes as controls.

2. Empirical analysis

We estimate the following OLS regression equation:

$$COHESION_{ij} = \alpha + \beta Final_j + \gamma EP7_j + \delta (FINAL_j.EP7_j) + \theta CONTROL_{ij} + \varepsilon_{ij},$$

where $COHESION_{ij}$ is a measure of the cohesion of party i on vote j, $FINAL_j$ is a dummy variable which takes the value 1 if a vote is a final legislative vote and 0 otherwise, and $EP7_j$ is a dummy variable which takes the value 1 if the vote is taken during EP7 and 0 otherwise. The Control variables include other factors that can potentially affect cohesion. The list includes policy area dummy variables as well as variables indicating identity of requesting political groups. The cohesion index is calculated as follows:

$$COHESION_{ij} = \frac{\max\{Y_{ij}, N_{ij}, A_{ij}\} - \frac{1}{2}\left[(Y_{ij} + N_{ij} + A_{ij}) - \max\{Y_{ij}, N_{ij}, A_{ij}\}\right]}{(Y_{ij} + N_{ij} + A_{ij})}$$

where Y_{ij} denotes the number of 'Yes' votes expressed by group i on a given vote j, N_{ij} the number of 'No' votes and A_{ij} the number of 'Abstain' votes.

This is a slightly modified version of the Rice Index adapted for legislatures were parliamentarians have three recorded voting options (Yes, No, and Abstain) rather than two (Yes, No). As one can see, maximal cohesion takes a value of 1 whereas it takes a value of 0 if votes within a party are equally divided between Yes, No, and Abstention.

The regression coefficient α measures cohesion on non-final votes (i.e. mainly amendments) in EP6. β measures the difference in cohesion between final legislative votes and other votes (in EP6). Coefficient γ measures the difference in cohesion between EP6 and EP7 (on amendment votes), and δ is the difference-in-differences estimator we are interested in.

Table 1. Difference-in-differences in Average Cohesion

	Pre-treatment (EP6)	Post-Treatment (EP7)	Difference Post-/Pre-Treatment
Treatment (final votes)	0.782***	0.793***	0.010
	(0.090)	(0.066)	(0.011)
	[62]	[151]	
Control (amendments)	0.760***	0.788***	0.028***
	(0.073)	(0.076)	(0.007)
	[384]	[198]	
Difference treatment-control	0.013	0.005	Difference-in-differences:
	(0.010)	(0.008)	0.016
			(0.012)

Note: *** p<0.01, ** p<0.05, * p<0.10. Robust standard errors in parentheses. Number of observations in square brackets.

In Table 1 we first present the basic difference-in-differences comparisons of the means of the four variables across the political groups. In the first column we show the average cohesion rate in final legislative votes and amendment votes in EP6 as well as their difference. Standard errors are indicated in parentheses and the number of observations indicated between

brackets. The second column gives the same calculations for EP7. The first row gives average cohesion on final votes in EP6 and EP7 as well as their difference. The second row does the same for amendment votes. The third column gives the difference between final legislative votes in EP7 and EP6, the difference between amendment votes in EP7 and EP6 as well as the difference between those two differences. The third row gives the difference between final and amendment votes in EP6, the difference between final and amendment votes in EP7 as well as the difference between these two differences. And the last element of the third column and the third row is the difference-in-differences and its standard error. We see that it is not significant. In other words, the simple difference-in-differences approach suggests that there is no significant difference in the cohesion levels of the political groups between requested RCVs and compulsory RCVs.

Table 1 indicates that average group cohesion, in particular on amendments, increased in EP7. While the increase in cohesion for amendments is small in magnitude, it is statistically significant. In addition, the statistics in Table 1 show that final votes are associated with higher cohesion, but the magnitudes of the effect are small and non significant.

Table 2. Difference-in-differences of Absolute Cohesion on Legislative Votes (clustered around voting blocs)

Variable	Final		EP7		Diff-in-Diffs		Constant		R-squared
	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	
Average p.group	0.022*	(0.011)	0.028*	(0.015)	-0.02	(0.017)	0.760***	(0.007)	0.039
ALDE	0.068***	(0.019)	0.01	(0.036)	0.04	(0.034)	0.853***	(0.020)	0.087
EFD (IND-DEM)	-0.048	(0.030)	-0.039	(0.035)	0.004	(0.041)	0.514***	(0.022)	0.020
EUL/NGL	-0.121***	(0.025)	-0.052**	(0.026)	0.032	(0.036)	0.920***	(0.010)	0.106
NA	0.105***	(0.033)	0.168	(0.03)	-0.173*	(0.041)	0.747***	(0.023)	0.097
EPP	0.039*	(0.022)	0.048	(0.045)	0.046	(0.044)	0.842***	(0.018)	0.093
S&D	0.069***	(0.014)	0.045	(0.033)	-0.006	(0.030)	0.874***	(0.018)	0.100
G/EFA	-0.018	(0.023)	0.056***	(0.019)	0.021	(0.019)	0.912***	(0.015)	0.072

Note: N=760. *** p<0.01, ** p<0.05, * p<0.10. Cluster (vote bloc) Robust standard errors in parentheses.

In Table 2 we show the difference-in-differences estimates for all political groups using clustered robust standard errors around voting blocks, where the latter are defined as a collection of votes on a particular legislative issue during a daily session. Votes belonging to a block may indeed share various similarities, such as the *rapporteur*, the timing, the substance of the issue, and so on. If legislators vote similarly on votes in a given block, there will be clustering around voting blocks, therefore the need to clustered standard errors around voting blocks. We have excluded two groups: Union for Europe of the Nations (UEN), a conservative group that existed in EP6 but not in EP7; and the European Conservatives and Reformists (ECR), which was formed at the start of EP7 when the British and Czech conservative MEPs left the European People's Party (EPP), but was not present in EP6. Note that because of this split, the composition of the EPP changed between these two sessions of the European Parliament. The EPP arguably became more ideologically homogeneous as a result. Similarly, the Europe of Freedom and Democracy (EFD, formerly Independence-Democracy in EP6) also changed its composition, as some of the former UEN members joined the new group in EP7. We say more later on whether composition effects had any effect.

Looking at the results, we see in Table 2 that difference-in-differences estimates are not significant for all political groups except for the non-attached MEPs.

Carrubba et al. noted that there was no proportionality between group size and the proportion of RCVs requested and also that certain committees were over- and underrepresented in RCVs relative to the general population of votes.[9] We hence control for these differences. In particular, this allows a *ceteris paribus* interpretation of our results. In Table 3 we enter controls

[9] Carrubba et al., 'Off the Record'.

for the legislative committee working on the relevant legislative proposal. As we can see, the difference-in-differences estimates are not significant.

Table 3. Difference-in-differences Regressions With Additional Control Variables

Variable	Average	ALDE	EFD	EUL/NGL	EPP	S&D	G/EFA
Final	0.021	0.072***	-0.041	-0.098***	0.026	0.058***	-0.030
	(0.013)	(0.021)	(0.034)	(0.029)	(0.020)	(0.017)	(0.022)
EP7	0.038***	0.031	-0.043	-0.051**	0.079***	0.030	0.062***
	(0.013)	(0.046)	(0.031)	(0.022)	(0.029)	(0.035)	(0.013)
Diff_in_Diff	-0.025	0.012	-0.005	0.050	0.005	-0.004	0.016
	(0.016)	(0.038)	(0.038)	(0.035)	(0.027)	(0.031)	(0.025)
Agriculture	0.021	-0.016	-0.012	-0.007	-0.020	-0.029	0.040*
	(0.029)	(0.028)	(0.070)	(0.048)	(0.029)	(0.035)	(0.020)
Transport	0.021	-0.016	-0.012	-0.007	-0.020	-0.029	0.040*
	(0.029)	(0.028)	(0.070)	(0.048)	(0.029)	(0.035)	(0.020)
Budget	0.016	0.050**	-0.006	-0.095	0.024	0.033**	-0.015
	(0.013)	(0.021)	(0.032)	(0.065)	(0.018)	(0.014)	(0.015)
Civil Liberty	-0.004	-0.096*	0.117**	0.000	0.003	0.009	0.001
	(0.013)	(0.056)	(0.047)	(0.030)	(0.022)	(0.020)	(0.023)
Constitutional Affairs	-0.001	-0.045	0.024	-0.172**	0.019	-0.005	0.003
	(0.022)	(0.030)	(0.047)	(0.079)	(0.019)	(0.027)	(0.015)
Development	-0.009	-0.006	0.073	-0.107	0.015	0.015*	0.034***
	(0.016)	(0.014)	(0.087)	(0.094)	(0.010)	(0.008)	(0.009)
Employment	0.019	-0.040	0.161***	0.028	-0.071	0.030*	-0.043
	(0.017)	(0.057)	(0.051)	(0.036)	(0.063)	(0.018)	(0.029)
Environment	-0.002	-0.053	0.017	0.033	0.001	-0.049*	0.007
	(0.011)	(0.032)	(0.031)	(0.023)	(0.023)	(0.028)	(0.015)
Fisheries	-0.014	-0.065*	-0.025	-0.040	0.015	-0.030	0.013
	(0.018)	(0.034)	(0.055)	(0.035)	(0.030)	(0.020)	(0.023)
Gender Equality	-0.088***	-0.146***	-0.096***	0.097***	-0.331***	-0.007	-0.076***
	(0.013)	(0.034)	(0.028)	(0.025)	(0.039)	(0.024)	(0.016)
Industry	-0.018	-0.071*	0.032	-0.188***	-0.010	-0.014	0.022
	(0.013)	(0.041)	(0.070)	(0.056)	(0.035)	(0.022)	(0.014)
International Trade	-0.006	-0.023	0.034	-0.154**	-0.008	0.018**	0.017
	(0.012)	(0.025)	(0.033)	(0.064)	(0.029)	(0.008)	(0.017)
Constant	0.761***	0.890***	0.490***	0.905***	0.868***	0.902***	0.919***
	(0.008)	(0.022)	(0.027)	(0.018)	(0.019)	(0.016)	(0.014)
R-squared	0.101	0.139	0.084	0.196	0.293	0.139	0.124

Note: Dependent variable = political group cohesion score. N=757.. *** p<0.01, ** p<0.05, * p<0.10. Robust standard errors in parentheses.

In Table 4 we further refine the controls by looking at the effect on party *j* of party *i* requesting a RCV, to take into account effects of RCV requests on parties not requesting the roll call. Since their voting behaviour may depend on who requested the roll call, we looked at the matrix of RCV requests on other parties. Committee controls are included but not reported in the Table. The main finding here, again, is that the difference-in-differences estimate is not significant. We do see some effects of requests of some groups on the cohesion of other groups, but the signs are both negative and positive.

Table 4. Controlling for RCV Requests by Political Group

Variable	Average	EPP	S&D	ALDE	G/EFA	EUL/NGL	EFD
Final	0.022*	0.098***	0.059**	0.078***	-0.006	-0.113***	-0.019
	(0.011)	(0.027)	(0.023)	(0.023)	(0.022)	(0.026)	(0.031)
EP7	0.028*	0.063*	0.071***	0.037	0.057***	-0.080***	-0.054*
	(0.015)	(0.035)	(0.027)	(0.029)	(0.022)	(0.027)	(0.032)
Diff-in-Diff	-0.016	-0.040	-0.064*	-0.019	-0.001	0.082*	-0.016
	(0.017)	(0.038)	(0.033)	(0.036)	(0.029)	(0.042)	(0.048)
Request by EPP		-0.101**	-0.007	-0.009	0.007	0.047**	-0.075**
		(0.046)	(0.024)	(0.032)	(0.024)	(0.020)	(0.036)
Request by S&D		-0.055	-0.141***	-0.156***	-0.031	0.063*	0.078*
		(0.037)	(0.054)	(0.048)	(0.041)	(0.034)	(0.047)
Request by ALDE		-0.041	-0.017	-0.008	-0.034	0.053**	-0.029
		(0.038)	(0.029)	(0.045)	(0.040)	(0.027)	(0.054)
Request by G/EFA		0.024	-0.048***	-0.025	0.025*	0.042**	-0.008
		(0.036)	(0.017)	(0.031)	(0.015)	(0.020)	(0.037)
Request by UEN		-0.066	-0.099***	-0.030	0.000	0.040	-0.045
		(0.067)	(0.030)	(0.038)	(0.019)	(0.029)	(0.039)
Request by EUL/NGL		0.029	-0.056	0.012	-0.061	0.071***	0.065**
		(0.033)	(0.045)	(0.026)	(0.043)	(0.021)	(0.030)
Request by ECR		-0.142	-0.009	-0.031	-0.015	0.116***	-0.066
		(0.094)	(0.028)	(0.039)	(0.023)	(0.032)	(0.060)
Request by EFD		-0.017	-0.025	-0.043	-0.081**	-0.107***	0.015
		(0.031)	(0.030)	(0.034)	(0.039)	(0.038)	(0.044)
R-squared	0.039	0.178	0.196	0.155	0.162	0.161	0.059

Note: Dependent variable = political group cohesion score. N=760. *** p<0.01, ** p<0.05, * p<0.10. Robust standard errors in parentheses.

In Table 5, we compare our results with those of Mühlböck and Yordanova (M-Y). There are five differences between our analysis and theirs. First, the sample sizes are slightly different.

Our dataset includes 759 final legislative votes including single legislative votes. That is about 60 more final legislative votes, than the sample used by M-Y, which has 701 observations. We show that this difference plays no substantial role. Second, M-Y used weights. While using weights can be a good idea, it has some important consequences. For example, because they needed at least two votes to calculate their weights, they excluded all single legislative votes. As a result, their sample is further reduced: 530 rather than 701. Despite this drop in the sample size, we find that this alone has no effect on the sign or the significance of the main variable of interest, i.e. the interaction between final legislative votes and the EP7 dummy variable. But when pooling the data, the change in the sample size has an important effect. Third, M-Y used a non-linear method (GLM), whereas we use a simple linear method (OLS). The difference in method does not matter if we do not pool the data or use weights. Fourth, we focus on all political groups, whereas M-Y look only at the four main political groups (EPP, PES, ALDE, and Greens/EFA) that are the most cohesive political groups in the EP. This is an important difference that can explain the difference between our results and theirs. Finally, we look at the average cohesion of political groups. M-Y, however, pooled the data and ran a pooled analysis. It is not surprising then that the sample size they report is four times larger than the number of final legislative votes. Let us examine these differences one by one.

Column 1 of Table 5 replicates M-Y's fractional logistic regression or GLM method with weights assigned to data (Model 1 reported in their Table 3). They use pooled data for the four main party groups. The difference in difference coefficient is significant at the 10% level. Column 2 shows what happens when we use a linear model on the same data. The diff-in diff coefficient is not significant. Column 3 replicates our main result using our data (HNR). Column 4 uses their GLM method with weights to estimate their model but with our data. The diff-in diff coefficient is not significant. Columns 5 and 6 are similar to columns 3 and 4 except that only the

Table 5. Comparison with Mühlböck and Yordanova

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Pooled M-Y 4 EPG GLM	Pooled M-Y 4 EPG OLS	Simple average HNR All EPG OLS	Simple average HNR All EPG GLM	Simple average HNR 4 EPG OLS	Simple average HNR 4 EPG GLM	Simple average HNR 4 EPG OLS	Simple average HNR 4 EPG GLM	Pooled HNR 4 EPG OLS	Pooled HNR 4 EPG GLM
EP7	0.581***	0.0567***	0.0279***	0.160	0.0398***	0.414	0.0517***	0.545	0.0528***	0.379**
	(0.194)	(0.00989)	(0.00660)	(0.213)	(0.00813)	(0.295)	(0.0115)	(0.388)	(0.0136)	(0.157)
Final	0.512***	0.0512***	0.0217*	0.123	0.0395***	0.408	0.0395	0.393	0.0754**	0.578***
	(0.185)	(0.0187)	(0.0121)	(0.334)	(0.0147)	(0.474)	(0.0285)	(0.378)	(0.0323)	(0.168)
Diff-in Diff	0.824*	0.00900	-0.0163	-0.0901	0.0252	0.923	0.0163	0.794	0.0366	1.092***
	(0.430)	(0.0199)	(0.0143)	(0.425)	(0.0165)	(0.742)	(0.0299)	(0.827)	(0.0333)	(0.375)
Constant	1.813***	0.860***	0.760***	1.152***	0.870***	1.903***	0.866***	1.864***	0.805***	1.417***
	(0.106)	(0.00772)	(0.00391)	(0.125)	(0.00502)	(0.159)	(0.00878)	(0.206)	(0.0114)	(0.0885)
No Obs.	2,120	2,120	759	759	759	759	530	530	2120	2120
R-squared		0.072	0.039		0.167		0.141		0.083	

Note: *** p<0.01, ** p<0.05, * p<0.10. Robust standard errors in parentheses.

four main party groups are included. Columns 7 and 8 differ from columns 5 and 6 only in the fact that we excluded our 58 extra votes. In columns 5-8, the diff-in diff coefficient is not significant. We then performed analysis on pooled data, keeping the sample size identical to M-Y (that is, we dropped the 60 extra votes per political group that we have in our sample so that we analyze exactly the same votes). In column 9, we use pooled data with OLS and in column 10, we use GLM. As can be seen, the OLS coefficient is not significant, but the GLM coefficient is positive at the 1% level. We also looked at whether using weights has an impact. Using pooled data from our sample, while keeping the sample size identical, we ran the analysis with and without weights. Keeping everything else unchanged, using weights or not did not change the results: with GLM, in both cases the main variable came out as positively significant (0.82 p-value=0.055 and 0.87 p-value=0.058), whereas with OLS both methods gave a positive but insignificant result (0.009 p-value=0.65 and 0.015 p-value=0.25). Although using weights per se did not change the results, it affects the estimates in the pooled analysis because it substantially reduces the sample size. With weights, the sample size was 2120 but without them the sample increased by 31% to 2804. Given that this is a substantial (31%) increase, not surprisingly as consequence the estimates became more significant.

To conclude, we see that a significant diff-in diff coefficient can be obtained if we use jointly GLM with pooled data on only four party groups. If we drop one of these conditions, the coefficient ceases to be significant. The use of pooled data produces larger samples, which gives precision, but it tends to overlook serial correlations, i.e. the iid assumption becomes problematic. Note that if we use pooled data for all party groups (not included in Table 5), the diff in diff coefficient is negative and insignificant. Also if we just discard EPP from the analysis, then the main variable of interest becomes insignificant. EPP became more cohesive in EP7 because its composition changed. The British conservatives decided to form a separate political group in the

(ECR European Conservatives and Reformists). Table 5 thus shows the difference in methods and results between M-Y and ourselves. We conclude from that comparison that our analysis is complementary to M-Y's. They look at the four main political groups and conclude that there is positive selection effect, while we look at all political groups and conclude that there is no selection effect. Both analyses exclude a negative selection effects.

The remark on the difference in composition of the EPP led us to control directly for a possible composition effect in the change of MEPs between EP6 and EP7. There was indeed a high turnover of MEPs between EP6 and EP7. Suppose that mandatory RCVs lead to lower cohesion. If the newly elected members adopt particularly cohesive behaviour, this would undo any decrease in reduced cohesion resulting from mandatory RCVs. To eliminate such possible composition effects, we looked at the voting behaviour of only those MEPs who served in both EP6 and EP7. For each MEP, we look at whether he or she voted with the group majority in a given vote. If the MEP voted with the group majority we counted this as 'loyal' voting, otherwise we counted his or her vote as 'disloyal' voting. Then, taking the ratio of loyal votes over all group incumbents' voting provides us with an overall 'loyalty score'. These loyalty scores should be highly correlated with our cohesion index.

We computed this measure of loyalty for both incumbent MEPs (in both EP6 and EP7) and new MEPs (in EP7). The results using only the incumbents should give us a strong indication of the effect of the new voting rule for those who served in both parliaments. Table 6 presents the difference-in-differences estimation on the loyalty score of MEPs who were in EP6 and EP7, aggregated by political group. We again looked at the difference between final legislative votes and other legislative votes in EP6 and in EP7. As can be seen from Table 6, the difference-in-differences estimate is nowhere significant. We thus conclude that the absence of

significant difference in voting cohesion between mandatory and requested RCVs cannot be explained by a composition effect.

Table 6. Loyalty Score Regressions

Variable	L-EPP	L-S&D	L-ALDE	L-G/EFA	L-EUL/NGL	L-IND-DEM	L-EFD
final	0.029***	0.023*	0.025**	0.028*	0.016	-0.001	0.032**
	(0.008)	(0.013)	(0.012)	(0.015)	(0.015)	(0.015)	(0.016)
EP7	0.053***	0.004	0.014	0.043**	0.023	-	0.034
	(0.016)	(0.018)	(0.019)	(0.019)	(0.014)	-	(0.025)
Diff-in-Diff	0.003	0.019	0.003	0.015	0.010	-	0.022
	(0.015)	(0.020)	(0.021)	(0.019)	(0.019)	-	(0.026)
Constant	0.864***	0.880***	0.873***	0.893***	0.869***	0.944***	0.888***
	(0.008)	(0.010)	(0.009)	(0.014)	(0.009)	(0.010)	(0.014)
Observations	760	760	760	760	760	408	759
R-squared	0.219	0.065	0.039	0.126	0.035	0.000	0.090

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

3. Conclusion

Good theoretical and empirical arguments have been put forward to claim that there should be a strategic bias in the selection of RCVs in the European Parliament. In this note, we used a difference-in-differences design to estimate the extent of strategic bias in roll call votes in the European Parliament. We exploited a new rule introduced in EP7, whereby roll call votes became mandatory on final legislative votes. Results from our estimation suggest that there is no significant effect on the voting cohesion of the political groups in the European Parliament when RCVs are mandatory as opposed to when then are requested (strategically) by a political group. This is true for all the political groups represented in the European Parliament. These results are also robust to the introduction of various controls, such as the type of issues voted on, the identity of the political group requesting a roll call vote, or various definitions of the control group. Moreover, we control for composition effects, to take into account changes in the individual

membership of the European Parliament between EP6 and EP7. These results suggest that in practice, despite strategic motives in RCV selection, cohesion is not significantly different from RCVs where the strategic selection motive is absent.

Whilst our analysis shows that there is no significant difference in party cohesion between mandated and requested roll call votes, there may still be a difference in cohesion between roll call votes and secret votes because of the effect of transparency on legislative behaviour. And these transparency effects might explain the results of Hug, rather than the strategic request of roll call votes.[10] When voting is non-transparent (in secret), there are many more possibilities for collusion between representatives and special interest groups, opportunities for corruption, and so on. Obviously, further research is needed to estimate the magnitude of selection effects due to transparency effects.

[10] Hug, 'Selection Effects in Roll Call Votes'.