Explaining Political Selection: What Factors Determine One’s Party-List Rank at t+1? (Bachelor’s Thesis in Statistics)

Michal Smrek
Supervisor: Prof. Johan Lyhagen
Uppsala University, Department of Statistics
VT2014

June 5, 2014
I. Abstract

This thesis contributes to the under-researched field of political selection, namely one’s re-selection onto the party list after one has been elected in the previous term. The theoretical rationale is to introduce a broader concept of political selection to a field mostly focused on political recruitment, one’s first point of entry into Politics. We show that the framework developed to study political recruitment can be adapted to study any kind of political selection that involves a broad pool of aspirants from which successful candidates must be selected. To this end, we utilise a panel dataset containing data on 387 Czech legislators covering the period between 1996 and 2013. Using fixed- and random-effects panel models, we show that voting along the party line and preferential vote share at time $t$ are strong predictors of getting a better party-list rank at $t+1$. Legislative experience, however, is negatively associated with how well one fares at the re-selection process. We also provide evidence that it is left-wing parties rather than their right-wing counterparts that discriminate against women at the re-selection stage. The study thus contributes, directly or otherwise, to debates on women’s representation, political careers and re-election.

**Key words:** re-election, fixed effects, random effects, panel data.
## I. Contents

1 Abstract 2

2 Introduction 4

3 Theory 5
   3.1 Political Recruitment ........................................ 5
   3.2 Political Selection ............................................. 8
   3.3 Why Re-Selection Matters? ..................................... 11
   3.4 Summary ..................................................... 12

4 Hypotheses 13

5 Case Selection 14

6 Data and Variable Description 16
   6.1 Dataset ....................................................... 16
   6.2 Variable Description ......................................... 18

7 Method and Model Specification 20

8 Descriptive Statistics 21

9 Empirical Analysis 23
   9.1 Full Random-effects Model ...................................... 23
   9.2 Restricted Models ............................................... 25
   9.3 Assumptions testing ........................................... 28
      9.3.1 Influential observations and outliers .................. 28
      9.3.2 Normality of the error terms ............................. 29
      9.3.3 Homoscedasticity ......................................... 30
      9.3.4 Multicollinearity ......................................... 31
      9.3.5 Linearity ................................................ 31
   9.4 Alternative model specifications ............................. 33

10 Discussion 35

11 Conclusion 37

12 Bibliography 38
   12.1 Theoretical Framework ........................................ 38
   12.2 Methodological Design ....................................... 40
   12.3 Other Online Resources ..................................... 40

13 Appendix 41
II. Introduction

While the issue of political recruitment receives wide attention by political scientists, very little research is done on the factors that determine one’s success in other types of political selection (Norris, 1997; Krook, 2010; Rahat, Hazan, 2001; Rahat, 2001; Bjarnegård, 2013, Kenny, 2013; Siavelis, Morgenstern, 2008; Norris, Lovenduski, 1995). While some studies implicitly conflate different selections (Norris, Lovenduski, 1995), they are inadvertently different to a significant extent, meaning that some of the factors that matter for one’s selection onto party lists, for instance, are very different from those that matter for one’s political recruitment. This study argues that political recruitment is merely a subcategory of a broader family of political selection. An attempt is made to pave the way for a theoretical framework on political selection by investigating the factors that determine one’s re-selection onto the party list after one has been elected in the previous term – another subcategory of political selection. In proportional representation systems, one’s position on the party list to a large extent determines whether one will be elected or not. This is why, selection on the party lists is interesting from a scholarly point of view and has a potential to contribute to debates on political careers, political turnover and studies on re-election.

This bachelor’s thesis examines the factors that determine what ranking on the party lists political candidates receive. By focusing on what determines one’s placing on the party list as a proxy for how well one fared at the re-selection process, this study overcomes the fundamental problem political-selection literature faces of not being able to observe the unsuccessful aspirants: those aspirants who do not get selected (Norris, Lovenduski, 1995). At the same time, however, it falls closely within the umbrella of what political-recruitment and political-career literature aims to examine, namely the factors that determine one’s political selection (Siavelis, Morgenstern, 2008). By focusing on party-list ranking, we also have an opportunity to, among others, investigate the often-debated hypothesis that party recruiters intentionally hinder women’s chances to be re-elected and reach senior positions (Norris, Lovenduski, 1995; Bjarnegård, 2013, Kenny, 2013).

Before we move on with theoretical review and present our main argument, some definitions are in order. In order to avoid conceptual confusion, we distinguish between political recruitment as the first step towards one’s parliamentary career (Norris, 1997; Siavelis, Morgenstern, 2008) and political selection as an umbrella concept that encompasses all kinds of political selections. Apart from political recruitment, other types of political selection would be any subsequent selection for a political office (executive office, senior parliamentary office). Given that it is political recruitment that has received most scholarly attention to date, we review the theoretical frameworks developed to study political recruitment and try to adapt them to our needs (Norris, Lovenduski, 1995; Fleischer, Seyfriend, 2013; Back, Dumont, Meier, Persson, Vernby, 2009).

For the purposes of empirical investigation, we make use of a panel dataset on
Czech legislators which was put together by the author. This rich dataset allows us to use various statistical techniques, predominantly fixed- and random-effects panel data regression analysis, to investigate our hypotheses. We find that voting along the party line and share of preferential votes are good predictors of receiving a better spot on the party list at t+1. We also find a somewhat counterintuitive result that political experience is negatively associated with one’s party-list rank at t+1. We attribute this finding to the natural party turnover and the fear party gatekeepers might have that the more experienced legislators will pose challenge to their leadership.

What follows is a theoretical chapter where we introduce our theoretical framework. Next, we present our hypotheses, dataset and variables and introduce the models that will be used to investigate our hypotheses. We then proceed with an empirical analysis which is concluded by a discussion of results. Finally, we conclude.

III. Theory

3.1 Political Recruitment

While the issue of candidate recruitment is attracting growing scholarly attention, very few scholars tend to view recruitment as a part of a broader family of political selection and often fail to analyse what happens after one has been initially recruited (Norris, 1997; Krook, 2010; Rahat, Hazan, 2001; Rahat, 2001; Norris, Lovenduski, 1995). Yet, as already stated, in many political systems one faces a new way of recruitment on a regular basis, e.g. when considered for committee or executive positions, or when being placed on the party list (Fleischer, Seyfried, 2013; Siavelis, Morgenstern, 2008; Back, Dumont, Meier, Persson, Vernby, 2009). This provides political parties with considerable discretionary power in deciding who will stay on and who will have to leave Politics (Siavelis, Morgenstern, 2008).

Recruitment and the broader political selection are so closely intertwined that we view the former as a subcategory of the latter. Both phenomena involve some sort of a pool of aspirants and certain screening mechanism which aids the selection procedure (Siavelis, Morgenstern, 2008). During the process, each aspirant’s suitability is carefully scrutinised with the aim of selecting an individual most suited for the position in question (Norris, Lovenduski, 1995). What inevitably distinguishes the two concepts (and ultimately distinguishes all subcategories of political selection as we will argue below), however, is the specific list of formal and/or informal screening criteria which are considered important for each of the methods of selecting candidates (Fleischer, Seyfried, 2013). While one might reasonably expect some overlaps between the two lists of different factors, some will naturally be unique for either recruitment or selection. Siavelis and Morgenstern (2008) provide a meaningful theoretical framework for distinguishing between recruitment and selection. According to their view, candidate
recruitment is a mechanism through which potential candidates come to be recruited into Politics and political selection as the process in which candidates come to be selected from a certain pool. They further note that: “Once candidates are victorious, the rules regarding re-election and whether candidates have static or progressive ambition continue to affect the extent to which they have to think about re-election and reentering of the recruitment arena, which in turn will affect their decisions about policy and behaviour and about support for the executive” (Siavelis, Morgenstern, 2008). What this means in real terms is that one can reasonably regard the selection prior to one’s re-election, one’s selection for positions of responsibility, one’s selection for executive positions, etc. as brand new political selections. A new pool is formed from which a final candidate(s) needs to be drawn. In case of re-selection onto the party list, for instance, each candidate is yet again scrutinised by the gatekeepers and either placed on a more or less electable spot or not placed on the party list at all.

Given that it is political recruitment that receives scholarly attention, the broader concept of political selection is poorly theorised (for a few examples, see Fleischer, Seyfried, 2013; Back, Dumont, Meier, Persson, Vernby, 2009). Yet, given that the former is a subcategory of the latter, the two inevitably have a lot in common. In order to build a theoretical framework for political selection then, we review the important theoretical work done on political recruitment.

The supply and demand framework of candidate recruitment proposed by Norris and Lovenduski (1995) still dominates and gives structure to debates within the field of political recruitment (Krook, 2010; Bjarneård, 2013, Kenny, 2013; Norris, 1997). According to this view, political recruitment is governed by a free-market-like mechanism of supply and demand. On the demand side, one finds party recruiters who screen potential candidates – or the so-called aspirants, according to some criteria which party recruiters deem important for the position in question. These include formal qualifications (such as education), speaking abilities, age, financial resources, political connections, name recognition, group networks, organisational skills, ambition for office, incumbency status, loyalty potential, etc (Norris, 1997). The supply side is defined by the ability and motivation of prospective aspirants to come forward and ask for consideration. This broadly includes potential aspirants’ resources (financial endowment, type of employment and career flexibility, politically relevant skills) and motivation (one’s willingness to step forward). Both supply and demand interact in order to produce an outcome of interest – political candidates. A candidate is recruited when a supply-demand equilibrium is reached.

Though intuitively convincing, the formal supply and demand framework came to be criticised on various fronts. Krook (2010), for instance, argues that the dynamics of supply and demand come to be distorted by gendered norms and practices. When asking herself why too few women step forward, Krook blames traditional socialisation patterns which lead women to believe that their place is in the family rather than Politics. On the demand side then, gendered social
norms shape who is qualified and who is desirable. Party recruiters who are not insulated from the prevailing social norms, intentionally or otherwise, design such formal and informal recruitment criteria which hinder women’s chances to be recruited. Other scholars join the critique and argue that it is not only women who are subject to discriminatory practices. According to Norris, other potential sources of discrimination are occupation, class, race, etc. Kenny (2013) joins Krook in her critique. In her view, both supply and demand are inherently distorted. In her view, both formal and informal rules of recruitment are influenced by dominant norms which favour middle-class, College-educated middle-aged men.

On the demand side, party recruiters engage in direct or imputed discrimination, while on the supply side, the prevailing norms have an effect on potential aspirants’ self-evaluation of their own merits and suitability for political office. As a result, potential aspirants undervalue their own merits and perceive themselves less qualified for elected office. As a matter of consequence, they fail to step forward even if they have the necessary resources and motivation simply because they believe they would stand no chance at the recruitment stage.

Despite these criticisms, the supply-demand framework is a powerful theoretical tool for understanding candidate recruitment. While social norms are not easy to measure, they nevertheless constitute an inherent part of both the supply and demand. While their influence on both is unquestionable, this critique does not prevent us from using the supply-demand model as a tool for understanding political recruitment. There is, however, a somewhat complementary model explaining candidate recruitment which overcomes the aforementioned problems associated with the supply-demand framework. The principal-agent framework views party recruiters as principals who are selecting party members, agents (Fleischer, Seyfried, 2013; Back, Dumont, Meier, Persson, Vernby, 2009). The fundamental problem they face is that once selected, agents will become principals in their own right in some areas (Fleischer, Seyfried, 2013). If they become regional party officials, they are responsible for representing the party within the region, among others. If elected to national or regional legislatures, they can independently vote on legislation. In order to maximise the probability that aspirants will further party interests after they are selected (and after they have become principals in their own right), political parties need to put an elaborate screening mechanism in place. Given that political parties cannot evaluate aspirants’ political qualities (assuming that they have not previously worked as politicians or party officials), they need to use other proxies to maximise the probability that they will act the way political parties desire them to act. To this end, they look at aspirants’ previous profession, educational level attained, managerial and public-speaking skills, etc. All these indirect indicators are used to assess whether the aspirant under evaluation could be expected to carry out his/her duties and further his/her party’s interests (Fleischer, Seyfried, 2013). The theory goes that in order not to take any risks, principals tend to recruit agents who are similar to them. This is why, middle-aged, College-educated, middle-class men tend to be overrepresented among the new recruits.
The inherent problem associated with research on political recruitment is that one rarely gets an access to the whole pool of aspirants. The unsuccessful ones can rarely be observed and if their identity is known, they might not be willing to take part in a scientific study. As a result, recruitment literature is limited to a few party-level studies, which suffer from selection bias and missing observations. Moreover, such studies are usually qualitative or theoretical in their nature, which obviously limits their external validity. This study thus tries to overcome the problem of missing observations and explore political recruitment from a quantitative perspective.

Another problem associated with recruitment research is the complicated nature of the supply-side which is, to some degree based on self-assessment of one’s merits. This poses great difficulties when trying to identify who the potential aspirants might be who are interested in seeking political office but keep undervaluing their own merits and do not step forward. The aforementioned principal-agent framework gets around this problem by focusing predominantly on the demand-side (what do the recruiters want?) and taking the pool of aspirants for granted (those who fail to step forward are not considered aspirants).

3.2 Political Selection

Given that political selection is poorly theorised in the literature, this theoretical framework borrows a lot from the recruitment literature. In the remainder of this chapter then, we build a theoretical framework on political selection by comparing and contrasting it to the concept of political recruitment. The focus is given on political re-selection onto the party list – another subcategory of political selection, which received virtually no attention in the literature to date. Political re-selection concerns those who have been elected in the previous term and are seeking re-election. Prior to being re-elected, they must be re-selected by their political parties.

Political selection can also be viewed through the lenses of the supply and demand framework discussed above. With regard to party’s demand, some of the variables which are important for one’s recruitment might not necessarily be that important for other types of selection. For example, one set of factors might be relevant for one’s repeated selection onto the party list which we dub re-selection and another set of factors is likely to affect the selection process of Prime-ministerial or ministerial candidates (Fleischer, Seyfried, 2013). This simply means that both the demand and supply change depending on the type of political selection. We can thus distinguish between political recruitment as the first entry to Politics, political selection onto the party, political re-selection onto the party list, selection into senior legislative positions (Committee chair, speaker), executive selection, etc (Fleischer, Seyfried, 2013; Back, Dumont, Meier, Persson, Vernby, 2009).

As already mentioned, focusing on the factors that determine one’s ranking
on the party list overcomes the fundamental problem of political recruitment research. This is because everybody who is selected onto the party list receives his/her rank. We thus do not have the problem of missing cases. Given that it is reasonable to expect that those who fared the best at the selection process receive the best spots on the party list, one's ranking can thus be seen as a good proxy for how one fared at the selection process. This way, both the selection process of first-time entrants and those who are re-selected can be studied. This study focuses on re-selection onto the party lists, that is determination of one's ranking on the party after one has already served the previous term in the legislature. The benefits of focusing on re-selection rather than first-time selection are discussed in the next subsection.

We now proceed by comparing and contrasting the recruitment model developed above to re-selection with an aim of creating a theoretical model which could help us to understand re-selection. We utilise both the supply-demand and agent-principal framework. While pre-political occupation is thought to be of crucial importance at the recruitment stage, its importance diminishes when it comes to one's re-selection onto the party list (Fleischer, Seyfried, 2013). This because previous occupation serves as a screening tool which helps party gatekeepers to check whether the aspirant in question possesses the practical skills necessary that will allow her to carry out legislator’s duties. Yet, after one has already been elected at least once, the importance of one's pre-political experience in re-selection is likely swiftly diminish. Instead, one’s political experience will play a dominant part. This might include one’s loyalty (voting records) and overall experience in Politics (number of years served in the party or the legislature). In their study of executive selection, Fleischer and Seyfried (2013) show that conventional factors which are deemed important for recruitment, such as education and other socio-demographic factors, do not seem to increase one’s chances to be selected. On the contrary, it is party-relevant experience and being a party insider that matter. Age, on the other hand, could potentially remain an important factor – with older politicians being less likely to be re-selected after they have passed certain threshold. The remaining formal criteria identified by Norris (1997), such as financial resources, speaking ability and formal qualifications (such as education) seem to intuitively matter more for initial recruitment, rather than re-selection. What seems to be a decisive factor, however, is party loyalty. Since re-election makes politicians more influential, parties need to make sure they do not select somebody who is likely to pursue ends which are not in their interest (Siavelis, Morgenstern, 2008).

The supply side is more complicated than the recruitment supply side. This is primarily because the motivation argument gets more multilayered. In the recruitment case, we only dealt with a dichotomous motivation structure: aspirants either desired to enter Politics or not. In the re-selection case, however, the picture gets more nuanced. We might encounter 4 different kinds of motivation: A. MPs might not be interested in pursuing political career any longer (negative motivation), B. they might wish to enter local Politics instead, perhaps believing
that they can be more influential there (regressive motivation), C. they might wish to get re-elected (static motivation), or, finally D. they might wish to leave Politics in order to seek higher electable or non-electable positions of power: progressive motivation Maestas, Fulton, Maisel, Stone, 2006; Hibbing, 1999; Samuels, 2000; Leoni, Pereira, Renno, 2004). This makes the picture more complicated. The resource argument works in a similar fashion as in the recruitment case. Existing political experience might, however, improve legislators’ position to better self-evaluate whether they have the necessary qualities to stay in Politics.

The aforementioned framework applies to the proper re-selection, that is we have a pool of aspirants who desire to be included on the party list again. By focusing on which position selected individuals receive on the party list after they have been re-selected, one overcomes the many methodological problems associated with the multilayered nature of the supply curve discussed above. Party-list ranking, moreover, is intuitively a good indicator of how well each individual fared at the re-selection stage. It does allows us to investigate the same phenomenon, while overcoming the problem of insufficient data on the supply side.

The already mentioned principal-agent framework appears to be useful for understanding re-selection and ranking determination, given that it is primarily demand-focused. Principals, or party gatekeepers, select their agents who will however become principals in their own right as far as legislative voting and public statements are concerned after they have been elected. When recruiting people, parties minimise the probability that the new recruits will not pursue the party line once elected by carefully scrutinising their previous skills, education, contacts, etc. When it comes to those who have already been elected and who the party gatekeepers know, there are better ways of evaluating whether they are likely to remain loyal to the party or not. If the elites want to maximise the probability that all selectees will uphold the party line once elected, they will most likely consider the following factors: membership in the party, legislative experience, and voting records. The longer one has been a member of the party at the time of re-selection, the higher should be one’s chances to be re-selected. Same logic applies to legislative experience. Finally, comparing individual voting records to the official party line could be a good indicator of one’s loyalty to the party line.

We have thus shown that the theoretical frameworks that were developed for investigating political recruitment are well-suited for other forms of political selection, too. We have tried to adopt the principal-agent framework to another subcategory of political selection – re-selection onto the party lists. By empirically investigating whether the factors which we suspect might determine one’s re-selection truly matter, we aim to justify the need for viewing political selection in broader theoretical terms, rather than conflating all kinds of political selection with political recruitment.
3.3 Why Re-Selection Matters?

Before we move on to present our hypotheses, we need to theoretically establish why re-selection and candidate ranking are interesting from a scholarly point of view. First of all, re-selection is a gateway to re-election, which is theoretically tied to such concepts as political influence, seniority, empowerment, substantive representation, etc (Young, 2000; Mansbridge, 1999; Phillips, 1995; Swers, 2005; Rosenthal, 1997; Carey, Niemi, Powell, 2000). In this final section, we review the scholarly views on the merits of re-election and thus attempt to justify the merits of the present study from a scholarly point of view.

As already mentioned, one’s ranking on the party list has a direct impact on whether the individual in question has a chance to get re-elected or not (Siavelis, Morgenstern, 2008). This holds especially in closed party-list systems, where voters have no say in reshuffling the party lists submitted by political parties. Even in most open systems, however, voters only seldom change the ranking of candidates on the party lists. The position one receives is thus closely tied to one’s re-election chances.

Re-election is at the centre of a number of debates. In the field of gender- and minority studies, re-election is closely tied to one’s capacity to engage in critical acts, or - in other words, capacity to meaningfully represent one’s group - substantive representation (Young, 2000; Celis, Childs, Kantola, Krook, 2008; Franceschet, Piscopo, 2008). This is because as the number of re-elections increases, so does the time the legislator in question remains in the legislature (Folke, Rickne, 2012). This leads to his/her name recognition both within the legislature and among the general public, directly empowering the legislator in question. Moreover, re-election allows one to stay on in the legislature for longer and thus gather more political experience (Fleischer, Seyfried, 2013; Folke, Rickne, 2012). Political experience is considered to be one of the key determinants of senior position allocation both within one’s party and the legislature at large (Ibid). Furthermore, by prolonging the time one spends in the legislature strengthens one’s party’s trust towards this individual, which makes them more likely to appoint her to various positions of power (Fleischer, Seyfried, 2013). The principal-agent field of research emphasises the need of political parties to maximise the probability that once in position of power their representative will not work against party’s interests (Bäck, Dumont, Meier, Persson, Vernby, 2009). Given that the power of parties to remove their nominees from various posts is limited after they have been appointed, they maximise this probability by carefully screening potential candidates and evaluate their loyalty, which is logically a function of how much they trust the individual in question. Finally, one could perceive re-election as some kind of an accountability measure through which voters express their (un)happiness with their representatives (Young, 2000; Celis, Childs, Kantola, Krook, 2008; Young, 2000). Hence, as the number of re-elections increases, so does the empowerment of the re-elect and the respect and autonomy he/she enjoys among his/her party colleagues. This also increases one’s visibility
and hence autonomy in preparing and pushing through various crucial bills. In summary then, the longer one stays in the legislature, the more discretion and autonomy one is endowed with. Finally, by prolonging one’s stay in the legislature, re-election also makes a given legislator more known and recognised. This recognition, both by voters and the media, might have a positive impact on how the legislator in question perceives her own autonomy and make her more likely to become more active.

The statement that re-election facilitates political influence is no assumption. There is a growing literature showing that re-election is indeed a strong prediction of one’s likelihood to reach influential positions within the legislature (Committee chair, speaker, etc.) or one’s likelihood to be considered for executive office (Reingold, 2006; Folke, Rickne, 2012; Fleischer, Seyfried, 2013).

It thus becomes clear that re-election is both an important and interesting topic of empirical investigation. It is closely tied to debates on political careers, legislative turnover, political influence and seniority, accountability, and substantive representation. It thus becomes clear that it is both interesting and relevant to investigate factors that determine one’s ranking on the party lists, which is directly connected to one’s re-electability.

3.4 Summary

We have introduced a number of key theories and the summary of the key theoretical framework is thus in order. As indicated, this thesis deals with re-selection onto the party list after one has been elected in the previous term. Given that one’s rank on the party list to a large degree determines one’s chances to be re-elected, we find our focus both theoretically and empirically interesting. Using the principal-agent theoretical framework, we have theorised that when recruiting, party principals (the gatekeepers) need to maximise the probability that those who get selected with further their party’s ends and not act against their party’s interests. To achieve this end, party recruiters need to put in place an elaborate screening mechanism, which allows them to minimise the probability of selecting an undesirable agent. At the recruitment and initial selection stage, party gatekeepers do not know anything about the future political performance of the selectees. This is why, they need to use a number of proxies, which, again, allow the recruiters to maximise the probability that the recruits will act the way the recruiters desire them to. When it comes to re-selection of those who have already been elected at least once onto the party lists, party gatekeepers have much better ways of assessing the aspirants’ suitability. This is because they know how the individual behaved in the legislature (whether he/she was a rebel or not) and can access this by looking at each individuals’ voting records. They also know how long the person spent in the legislature and how long he/she has been a party member. Additionally, they also know how many times the party has entrusted the individual with a position of responsibility (Committee chair,
parliamentary speaker). All these indicators are far superior measures of one’s loyalty and legislative performance than the secondary proxies party recruiters use at recruitment and initial selection stage.

Focusing on the party-list ranking as a proxy for how well each individual fared at the re-selection stage helps us to overcome the problem of missing data on unsuccessful aspirants.

IV. Hypotheses

Given these aforementioned theoretical framework, the set of hypotheses that shall be tested are intuitive:

**Hypothesis 1**: Those selectees who have voted along the party line receive a more electable (higher) place on the party list.

The rationale behind this hypothesis is that the more loyal the individual, the more likely it is that he/she will further his/her party’s ends if re-elected.

**Hypothesis 2**: The longer one stays in the legislature, the higher spot on the party list one receives.

This is because the longer one stayed in the legislature, the more time one has been successfully re-selected. Experience should thus be a fine predictor of one’s success at the re-selection stage.

**Hypothesis 3**: The longer one has been a member of the party at the time of re-selection, the higher spot on the party list one receives.

This hypothesis, although theoretically-relevant is not tested due to data availability (see below).

**Hypothesis 4**: The older one gets, the marginally less more likely one is to get a lower spot on the party list.

Hypothesis 4 is intuitive again. One might, however, expect that one’s age might be correlated with one’s experience, inducing collinearity. Moreover, the relationship between age and party-list rank will most likely not be linear. This is because the association between age and party-list rank could be positive at first up to certain threshold. After one has passed this threshold, an increase in one’s age will no longer matter that much or even become negatively associated with one’s party-list rank (U-shaped relationship). Due to the potential collinearity, age shall only be tested in the first model and will be left out from all subsequent models in favour of experience, which intuitively seems to be more interesting from a theoretical point of view.
Hypothesis 5: Women are expected to, on average, receive lower spots than men.

There is a great host of literature within the political selection discourse, which claims that party recruiters (who are mostly men) discriminate against female aspirants simply because of their gender. Female aspirants are perceived as less predictable and less trustworthy simply because party recruiters’ experience with female legislators is limited. Some also perceive women as being less fit for Politics in general, being more consensual and less imposing in their behaviour. If this is indeed the case, we expect that party recruiters might deliberately try to put women to lower party-list spots to prevent their re-election and eventual seniority.

Hypothesis 6: Holding a senior position within the legislature improves one’s chances to receive a higher spot on the party list.

Political parties only entrust senior positions (Committee chairs, speaker positions) to those party members they trust. Holding a position of responsibility could thus be seen as a good indicator of how trustworthy party seniors perceive each individual aspirant.

Hypothesis 7: One’s preferential vote-share at time t is positively associated with one’s position at party list at time t+1.

One’s preferential vote share (the share of the total party preferential vote individual i receives) is intuitively a strong predictor of one’s success at the re-selection stage. Party recruiters need to consider how well one fared in the previous elections, because this is one of the few accountability measures available at voters’ disposal. The higher one’s share of preferential votes, the more voters approve of this individual’s actions in the legislature.

All the aforementioned hypotheses shall be tested at the conventional, 5% level of significance. All the hypotheses shall be tested in one model. Should we not have enough evidence to reject or confirm some of the aforementioned hypotheses (e.g. get a non-significant coefficient), these shall be excluded from the model and a new model shall be estimated.

V. Case Selection

The (non-)availability of panel data on national legislators poses some restrictions in regards to what particular questions might be asked and having a non-data-driven case selection. Panel datasets on legislators are extremely rare: most commonly not available at all or subject to an ethical committee approval and a considerable fee (as is the case in Finland, Denmark, Sweden and Norway).
We have built a panel dataset comprising all Czech legislators elected between 1996 and 2013 (6 elections all in all). In order to build the dataset, we relied on public directories of the Czech Statistical Office, web portal of the Parliament of the Czech Republic, voting-record directory of the Czech Parliament and online resources of all the political parties which have or have had some parliamentary representation during the aforementioned period. The Czech Republic, however, was not only selected due to data availability.

The Czech Republic offers a good case on which our theory can be tested (Matland, Montgomery, 2003). Intuitively, our theory mostly applies to those countries where party gatekeepers have a major say in putting the party lists together and deciding on who gets selected and who does not. The literature distinguishes several methods of party list selection (for a good overview, see Siavelis, Morgenstern, 2008, p. 18, pp. 31-33). The selection can be centralised (party lists are drawn by national elites) or decentralised (party lists are decided upon by local party representatives). Furthermore, the mode of selection can be open (party primaries determine who gets a spot on the party list and/or which rank each candidate receives), or closed (senior party gatekeepers decide on who gets selected and which rank each recruit receives), bureaucratised (there are clearly codified rules of what should be taken into account at the re-selection stage) or patronage-based (the rules are blurred and majorly at the discretion of party elites). Intuitively then, party recruiters are most powerful in centralised, closed and patronage-based systems, which is where our theoretical framework should have the strongest explanatory power. This is because, in such systems party recruiters are the true gatekeepers - they decide on who gets selected and who does not. Czech parties, just like most of their counterparts in the region, rely exclusively on centralised, closed and patronage-based methods of candidate selection (see Matland, Montgomery, 2003). National party elites have major say in determining how the party list will look like and who will get on it. In light of this, the Czech Republic offers a most-likely case where our theory might hold.

Second, the theory is most likely hold in political systems where party loyalty is high. This is why, it will most likely hold in proportional-representation (PR) systems rather than majoritarian systems. In majoritarian systems, representatives owe their positions and thus loyalty to the voters rather than their parties. In PR-systems, representatives owe their allegiance primarily to their parties which expect much more loyalty in return as far as voting is concerned as in their majoritarian counterparts. This is especially the case in countries where more than 6 people are elected within each district (Siavelis, Morgenstern, 2008, p. 18). Siavelis and Morgenstern (2008) claim that if the number of representatives per district exceeds 6, these representatives are far less likely to feel loyalty to their local constituencies but rather their parties and representatives become national actors rather than regional actors. The Czech Republic is yet again a most likely case here. The magnitude in each of the 18 electoral district exceeds 6, in many by quite a wide margin (Parliament of the Czech Republic, 2014). This gives rise to a highly system with exceptional party loyalty, rather than regional loyalty.

15
Parties expect all their legislators to adhere to the party line, which provides a fertile ground for testing whether one’s adherence to party line has an effect on how well one fares at the re-selection stage.

Finally then, the theory’s explanatory power will differ based on the type of party list employed. Intuitively, the theory will mostly hold in closed-list systems where voters have no power to re-shuffle the party lists prepared by party gatekeepers. It will have weaker explanatory power in open party-list systems where voters, in addition to being able to re-shuffle the lists, may introduce new candidates onto the lists during elections. The Czech Republic lies somewhere in the middle with regard to this. Employing semi-open party lists, Czech voters are allowed to cast 4 preferential votes. If one’s preferential vote share exceeds 5%, this person moves up the party list. Even though Czech voters have this nominal power, very few exercise it. According to Matland and Montgomery (2003), Czech candidates only very rarely move up or down the party list. This is why, the Czech Republic can still be regarded a most-likely case, even as far as party-list type is concerned.

We have thus established that the Czech Republic serves as a most-likely case for our present theory. This means that should we establish empirical support for the theory, more testing will be needed in order to verify its cross-case validity. The study can thus be regarded as the first step towards building a theoretical framework explaining political selection.

VI. Data and Variable Description

6.1 Dataset

As already mentioned, a panel dataset has been built for the purposes of this thesis. The dataset comprises data on all the MPs who have been elected to the lower chamber of the Czech Parliament (House of Representatives) at least once and have been placed on the party list in the subsequent period (t+1) between 1996 and 2013. The dataset thus covers 6 elections to the Czech House of Representatives and contains 649 observations for 387 individuals.

The data has been built by the author of this thesis using publicly available directories of the Czech Statistical Office. Here, data on all candidates for each election are stored. The data are stored individually, however, and candidates have no unique ID numbers attached to them. We thus had to merge the different datasets for each election using candidates’ names. After the datasets had been merged, one had to check for namesakes (different individuals with the same name). This was done with the help of Stata. After all the namesakes have been sorted, individual ID numbers have been assigned to each individual in the dataset.
The initial dataset contained some of the necessary background variables: age of each individual, position on the party list, position on the party list after the election, result of the election (whether the individual has been elected or not), total number of preferential votes, the share of the total preferential votes the party has received in the district, and the electoral district. The biological sex of each individual was coded based on each individual’s surname (in Czech language, all female surnames have a unique suffix). The education of each individual was coded as a categorical variable based on the available academic titles. In the Czech Republic, academic titles are always written together with each candidate’s name on the party lists. This includes Bachelor-, Master- as well as academic degree titles.

Some of the variables included in the dataset were added from different sources. The experience variable (how many years has the individual in question been sitting in the national legislature at time t) had to be manually coded for each individual in the dataset. This is because some of the parliamentarians in the dataset have already had some legislative experience before 1996. To this end, the online directory of the Czech Parliament has been used. Same applies to the party member variable (for how long has the individual been a member of his/her current political party). The aforementioned directory, however, does not contain information on party membership. This is why, we had to go through every single individual in the dataset and obtain this information from the Internet (political parties’ directories, MPs’ websites and bios, etc.). Using this method, we have collected relevant data on 237 of the total 387 individuals. We have contacted all the political parties and asked them to provide us with the missing data, but none of them responded until the deadline for the present thesis. We are thus wary to use this particular variable for the potential omitted observation bias. It is well likely that it is the more prominent MPs who post their bios online. Such individuals might thus be overrepresented in the sample which might bias the results. The variable is thus not included in any of the models.

Furthermore, two variables on party line voting have been collected from publicly available voting directories. The first indicator indicates the degree of agreement in voting between the individual in question and the average party vote. The first indicator gives all the votes equal weight, while the second one assigns them weights based on the proportion of the majority vote. The benefits and pitfalls of both methods of quantifying party-line voting are discussed below.

Next, two dummy variables indicate whether each individual served as a Committee chair in the legislature during time period t and whether the individual held the position of a speaker or a vice-speaker of the parliament. These were coded manually based on the publicly available directories on the website of the Czech Parliament.

Finally then, the dataset contains various background variables: whether one’s party has been part of the coalition or opposition at t, party vote share at t,
number of mandates at $t$ and whether the party is left- or right-leaning.

The dataset thus contains a wealth of information on Czech parliamentarians over time and is one of the few panel dataset that contain this much information on national legislators. Even the detailed panel dataset on Swedish legislators available at Statistics Sweden does not contain a number of important background variables this dataset does. This means that the dataset could potentially be of interest to other researchers and shall be made publicly available shortly.

6.2 Variable Description

In this subsection, we briefly describe the variables that are included in at least one of the models. We begin with the dependent variable.

**Party List Rank at $t+1$:** indicates the position on the party list at time $t+1$ for each individual $i$. The higher one’s position on the list the better. Note, however, that the best position is represented by number 1. This means that positive associations in the models (increase in $X$ leads to an improvement in $Y$) will actually have negative signs and vice versa. The mean of this variable is 4.06 with 1 being the minimum value and 26 being the maximum value. The standard deviation is 4.17.

The other variables are follows (note that not all of these are used in our models):

**Sex:** time-invariant dummy for individual $i$’s biological sex used as a proxy for gender with 1 indicating female. There are 70 women (113 observations) and 317 men (537 observations) in the dataset.

**Education:** 4-category categorical variable for individual $i$’s level of education. 1 indicates no University education, 2 indicates Bachelor’s degree, 3 indicates Master’s degree and 4 indicates academic degree (Ph.D, docent, professor, etc.).

**Age:** numerical variable for age. The mean age is 46.2 with minimum value being 21 and maximum value being 72. The Standard deviation is 9.2.

**Preferential Votes:** numerical variable for the number of preferential votes received by individual $i$ at time $t$. The mean age is 4830 with minimum value being 367 and maximum value being 67848. The Standard deviation is 5798.

**Preferential Share:** numerical variable for the percentage share of the total preferential party votes received by individual $i$ at time $t$. The mean age is 4.46 with minimum value being 0.16 and maximum value being 30.1. The Standard deviation is 4.09.
Elected at t: dummy variable indicating whether the individual in question has been elected at time t (1) or not (0).

Elected at t+1: dummy variable indicating whether the individual in question has been elected at time t+1 (1) or not (0).

On Party List at t+1: dummy variable indicating whether the individual in question has been re-selected onto the party list at time t+1 (1) or not (0). Note that 0, however, groups together both those individuals who have quit voluntarily and those who were interested in re-selection but were unsuccessful. This is why this variable is unsuitable for investigating political re-selection.

Senior position: dummy variable indicating whether the individual i has acted as a chair of one or more parliamentary committees at time t or not.

Speaker: dummy variable indicating whether the individual i has acted as a speaker or vice-speaker of the House at time t (1) or not (0).

Party Line Voting: numerical variable indicating whether individual i has been a loyal party-line voter measured as a proportion of votes cast by individual i which had been in line with the party line at time t. Party line is measured as an average party vote. The mean value is 0.91 indicating that most MPs are faithful party line supporters. The minimum value is 0.68 with the maximum being 0.98. The standard deviation is 0.04. There two different variables for party-line voting, one measuring the absolute degree of agreement, that is all the voting is taken into account. The other variable tries to weigh the importance of each voting by looking at how big the majority was at the time of the vote. The rationale is that the more controversial and thus important the vote, the less likely the opposition is to vote for it. This might, however, not always be the case. Voting for having a longer break might attract a wide majority of supporters, yet this break might decide whether the chamber will or will not support the proposal which they currently debate on. As a matter of consequence (and given that the Pearson’s correlation between the two measures is 0.96), we use the former variable which assigns equal weight to all the voting. As a robustness check, we do re-estimate all the model using the latter variable but the results are largely identical, most likely due to the high degree of correlation between the two variables.

Experience: numerical variable measuring the number of years individual i has been sitting in the national legislature up to time t. The mean value is 2.9 years, with 0 years being the minimum value and 18 years being the maximum value. The standard deviation is 3.6 years.

Party List Rank at t: indicates the position on the party list at time t for each individual i. The lower one’s position the better. The mean of this variable is 3.92 with 1 being the minimum value and 29 being the maximum value. The standard deviation is 3.47.
In addition to these background variables, there is variables indicating party vote share at t, number of mandates the party has received. There is also an ID variable for individual i’s party.

VII. Method and Model Specification

In order to investigate our hypotheses, a panel dataset has been collected on all Czech MPs elected between 1996 and 2013 (6 elections in total). It is thus an unbalanced micro-panel (short time series). This limits the viability of estimating a dynamic model and we thus settle with a conventional fixed- and random-effects panel regression.

Panel-data analysis carries with it a number of advantages. First, given that we are following individuals over time, the sample size increases considerably, making some analyses more viable (Asteriou, Hall, 2011, p. 416). Moreover, some panel data control for all time-invariant variables making the models more insulated from omitted variable bias.

In all models, the primary response variable is one’s position at time t+1. This is causally linked to a host of independent variables measured at time t: that is, degree of agreement with individual i’s party line between t and t+1, age at t, experience at t, positions of responsibility at t, etc.

We shall begin our analysis by estimating a random effects model. This is because one of our hypotheses is related to the effect of biological sex on one’s rank on the party list. Gender is time-invariant and it would be perfectly collinear with individual fixed effects, should we attempt to impose a fixed-effects model on it. The software would thus not estimate the effect of sex on the primary dependent variable of interest. This is why, we have to rely on a random-effects model, which comes at a cost. A random-effects model does not control for any omitted time-invariant factors, making the model more susceptible to omitted variable bias. Moreover, the model by construction assumes no correlation between the error term and the regressors, which, in our case, is an unreasonable assumption to make. Given the complexity of social phenomenon we are dealing with, it is almost certain that we are omitting some important predictors (some of them simply unmeasurable: such as strategic small-talk, etc.). The effects of all these omitted variables shall be contained in the error term. Provided that these unobserved variables might be correlated with the regressors, the aforementioned assumption shall be violated and the point estimates shall be biased. In the present situation, this is the best we can get, however.

After having run the random-effects model, we shall run Breusch-Pagan Lagrange multiplier (LM) test which helps us decide whether a random-effects model is needed or whether simple OLS is sufficient.
Provided that the only time-invariant variable we have in the model will be non-significant, we shall exclude it from the model, which will allow us to estimate both the random- and fixed-effects models. After having done so, we shall run a Hausman test which will help us to decide which model to keep.

After having settled on the final model, we shall check for all the assumptions we can: normality of the error term, homoscedasticity, no multicollinearity, linearity, no unusual or influential observations, etc. Testing for cross-sectional dependence shall not be necessary as we are dealing with a micro panel and this issue only affects macro panels with long time series (over 20-30 years).

To sum up then, the initial random-effects model is given as follows:

$$Y_{\text{PartyListRank}_{i,t}} = (a + v_i) + \beta_1 \text{Sex}_{i,t} + \beta_2 \text{LegExp}_{i,t} + \beta_3 \text{PrefVote}_{i,t} + \beta_4 \text{SeniorPosition}_{i,t} + \beta_5 \text{PartyLineVoting}_{i,t} + \beta_6 \text{Age}_{i,t} + u_{i,t}$$

where, $Y_{\text{PartyListRank}_{i,t}}$ is the response variable, $(a + v_i)$ is the variable constant for each section, with the rest being the aforementioned regressors and the error term. Note that in the actual model, SeniorPosition is measured by two variables.

This is a random-effects model that shall be modified to a fixed effects model, provided that the only time-invariant variable in the model shall turn out non-significant. The final fixed-effects model (before variable transformation) looks as follows:

$$Y_{\text{PartyListRank}_{i,t}} = a_i + \beta_1 \text{LegExp}_{i,t} + \beta_2 \text{PartyLineVoting}_{i,t} + \beta_3 \text{PrefVote}_{i,t} + u_{i,t}$$

where $a_i$ is the individual-specific intercept for each individual in the sample (fixed effects).

The software package Stata is used throughout the study.

VIII. DESCRIPTIVE STATISTICS

With most descriptive statistics being presented in earlier chapters, we only present a Pearson’s correlation matrix for the numerical and dummy variables and a Spearman’s correlation matrix where categorical variables are included, too. Figure 1 shows the Pearson’s correlation matrix (note that only those variables which shall be subsequently used are included).

As Figure 1 suggests, we have some weak and moderately strong correlations in our sample. The correlation between experience and voting along the party line, for instance, is moderately strong and negative. This means that the more
experienced individuals tend to be more rebellious in their voting. Same applies to speakers’ voting but to a much less significant extent. Unsurprisingly, there is a moderately strong correlation between experience and being a Committee chair or a speaker, which is intuitive. Moreover, the correlation between age and experience is moderate and positive, too, as expected. Normally, a correlation of this magnitude would not be likely to induce collinearity. One has to remember, however, that if the magnitude of all the correlations in the sample is relatively low, even weak correlations might induce collinearity.

We now look at the spearman correlation where education is included, too. We note that education is moderately positively correlated with preferential vote share, which might indicate the voters might be inclined to give their preferential votes to who they perceive to be experts. It is intuitively correlated with age, too. There is a negative correlation between education and party list rank,
which, bearing the way the variable is coded, is also intuitive. The more educated seem to receive better (in mathematical terms, lower) ranks on the party list.

As already mentioned, most correlations seem to be moderately weak or outright weak. This, however, does not mean that the models are unlikely to suffer from multicollinearity. On the contrary, provided that most correlations in a sample are relatively weak, even weak correlations might cause multicollinearity.

IX. Empirical Analysis

9.1 Full Random-effects Model

In this section, we present the results of our empirical analysis. We shall start by presenting a random-effects GLS regression with all the 7 independent variables specified by our 6 hypotheses (see above). The response variable is position on the party list at t+1 and the explanatory variables are Party-Line Voting, Preferential Vote Share, Age, Committee Chair, Speaker, Experience and Sex. As already mentioned, the Party Membership variable might induce bias by decreasing the sample size and causing overrepresentation of influential members of the legislature. Hypothesis 3 is thus not tested in this thesis. It is kept in the Hypotheses section, however, because it logically flows from the theoretical framework and should the data be available, it would have been tested.

As already mentioned, the weakness of a random-effects model is that we have to make a specific assumption that the variation across entities is random and that the random component is not correlated with the regressors (Torres-Reyna, 2011; Asteriou, Hall, 2011, p. 420; also see Baltagi, 1995, pp. 13-18). This is a very strong assumption to make and it is unlikely that this assumption holds in our case. The social phenomenon is likely driven by a number of informal factors, which are challenging to measure, such as trust. Our explanatory variables will be inevitably correlated with these omitted variables, which logically means that the random component will be correlated with the regressors. As a consequence, the estimates are likely to be biased and inconsistent (Asteriou, Hall, 2011, p. 420). We do, nevertheless, estimate a random-effects model, because this model allows us to include time-invariant variables, making random-effects the most suitable model we have at hand, despite its flaws. Moreover, the model has fewer parameters to estimate (we do not have person-specific dummies in the model), saving us degrees of freedom.

The estimates of the first model are presented below (Table 1). The full Stata output can be found in the Appendix (Figure 7).

Before interpreting the coefficients, we run the Breusch and Pagan Lagrangian multiplier test for random effects to decide whether a random effects model is appropriate or we should rely on a simple OLS model. Significant at 1% level of significance, the test clearly suggests that there is some variance across entities
Table 1: Full Random-effects GLM Model, Party List Rank at t+1 is the dependent variable

<table>
<thead>
<tr>
<th>Model:</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept:</td>
<td><strong>14.05</strong>* (4.53)</td>
</tr>
<tr>
<td>Party-Line Voting:</td>
<td><strong>-12</strong> (4.89)</td>
</tr>
<tr>
<td>Pref-Vote Share:</td>
<td><strong>-0.36</strong>* (0.037)</td>
</tr>
<tr>
<td>Age:</td>
<td><strong>0.053</strong>* (0.021)</td>
</tr>
<tr>
<td>Committee Chair:</td>
<td>-0.50 (0.57)</td>
</tr>
<tr>
<td>Speaker:</td>
<td>-0.24 (0.42)</td>
</tr>
<tr>
<td>Experience:</td>
<td>0.09* (0.05)</td>
</tr>
<tr>
<td>Sex:</td>
<td>0.49 (0.52)</td>
</tr>
<tr>
<td>$R^2$:</td>
<td>0.13</td>
</tr>
<tr>
<td>Observations:</td>
<td>649</td>
</tr>
<tr>
<td>Groups:</td>
<td>387</td>
</tr>
<tr>
<td>F-statistic:</td>
<td>124.40***</td>
</tr>
<tr>
<td>$\rho$:</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Notes: ** represents $0.01 \leq p < 0.05$, *** represents $p < 0.01$.
Heteroscedasticity-adjusted robust standard errors are in parentheses. All figures rounded to 2 decimal places.

and thus random effects are more appropriate than a simple OLS model (for the full Stata output, consult the Appendix, Figure 8). Passing onto the coefficients then, we note that some of the coefficients are statistically indistinguishable from 0. This means that we do not have enough evidence to settle whether the association between these particular variables and the response variable of interest is negative, positive or whether there is no association between the two. We thus do not have enough evidence to conclusively test Hypotheses 5 and 6.

Looking at the remaining coefficients, we note that as party-line voting of individual i increases by 1% at t, his/her party-list rank at t+1 improves by 0.12 places, holding all other variables constant. This means that no more than a 10% increase in party-line voting, increases one’s position on the party list by 1. We have thus found support for Hypothesis 1 at 5% level of significance: being a
disciplined party-line voters does seem to increase MPs’ position on the party list at t+1. One’s age is, as predicted, positively associated with party list rank at t+1. The older one gets, the worse position on the party list one receives, controlling for all other variables. 1 year increase in age decreases one’s position on the party list by 0.05 places. Hypothesis 4 is thus supported by the data. As predicted, the higher one’s share of preferential votes, the better one’s party-list rank. Hypothesis 7 is thus also supported by the data.

Finally, the longer one sits in the legislature, the worse position on the party list one gets, on average. This is counterintuitive because one would reasonably expect that the longer one stays in the legislature, the more trustworthy the individual in question becomes in the eyes of the gatekeepers. This, however, does not seem to be the case. One should also remember that the longer one stays in the legislature, the more recognised and influential the person becomes. Party gatekeepers might fear that very experienced individuals might pose challenge to their leadership and thus place them to less electable positions. This association is only significant at the 10% level of significance, however. Given that we have originally set the significance level at 5%, we conclude that we do not have sufficient evidence to reject or confirm this hypothesis. We do, however, have some counterintuitive hints.

The overall $R^2$ of the model is 0.13, which means that 13% of the variation in the dependent variable is explained by our regressors. Given the complex nature of the social phenomenon we are set to investigate and the fact that we are dealing with panel data, we find this amount of variation explained reasonably good. One should also remember that the $R^2$ is connected with the variation in the dependent variable. If we have a lot of variation, the $R^2$ will be high. If we do not have a lot of variation, the $R^2$ will be low by construction. Finally then, $\rho$ statistic shows how much of the total residual variance is due to differences between individuals. In our case, 40% of the residual variance is due to the differences between individuals, which, to some degree, justifies the use of random effects.

### 9.2 Restricted Models

Before we check whether the assumptions the model relies on have been fulfilled, we re-estimate the model leaving out the variables with non-significant coefficients.

We also leave out age despite the fact that the point estimate for this variable was significant. This is because experience is included in the model, too. Most commonly, age and experience are highly correlated with one another. This gives rise to collinearity which in turn inflates the estimates of standard errors (Forsberg, 2013). We suspect that collinearity between these two variables might have been present in our first model, too. Provided that the point estimate of the experience variable had a counterintuitive sign, we have decided to drop age and keep experience to further investigate the relationship between party list rank.
Table 2: Restricted Random-effects GLM Model, Party List Rank at t+1 is the dependent variable

<table>
<thead>
<tr>
<th>Model:</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept:</td>
<td>14.7***</td>
</tr>
<tr>
<td>(4.65)</td>
<td></td>
</tr>
<tr>
<td>Party-Line Voting:</td>
<td>-10.06**</td>
</tr>
<tr>
<td>(5.00)</td>
<td></td>
</tr>
<tr>
<td>Pref-Vote Share:</td>
<td>-0.35***</td>
</tr>
<tr>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td>Experience:</td>
<td>0.12**</td>
</tr>
<tr>
<td>(0.05)</td>
<td></td>
</tr>
</tbody>
</table>

\[ R^2: 0.11 \]
Observations: 649
Groups: 387
F-statistic: 98.94***
\[ \rho: 0.42 \]

Notes: ** represents \( 0.01 \leq p < 0.05 \), *** represents \( p < 0.01 \).
Heteroscedasticity-adjusted robust standard errors are in parentheses. All figures rounded to 2 decimal places.

at time t+1 and experience. Before doing so, we have taken a square of age to see whether the effect of age could be modelled as a parabola: positive effect on party-list rank up to a certain point and then a negative effect after this point has been reached (recall here that the lower value of the dependent variable the better). Indeed, the coefficient remained significant after this transformation and the overall model fit improved marginally. Given the moderately high correlation between age and various positions of responsibility in the legislature, we have at first dropped age but kept the two responsibility-related variables. Their coefficients remained non-significant, however, which means that they were not non-significant due to multicollinearity.

Table 2 shows the results of the restricted random-effects model (full output in the Appendix, Figure 9).

The second model, again significant at 1% level of significance, confirms the results yielded by the initial model. The party-line voting coefficient has decreased marginally in magnitude (from -12 to -10.06) but remained significant. The experience coefficient is now significant at the 5% level of significance and remained positive. This means that the longer one stays in the legislature, the worse spot on the party list one receives, on average. One year increase in experience leads
Table 3: Restricted Random-effects Vs. Fixed-effects GLM Models, Party List
Rank at t+1 is the dependent variable

<table>
<thead>
<tr>
<th>Model:</th>
<th>RE</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept:</td>
<td>14.7***</td>
<td>15.39**</td>
</tr>
<tr>
<td></td>
<td>(4.65)</td>
<td>(7.89)</td>
</tr>
<tr>
<td>Party-Line Voting:</td>
<td>-10.06**</td>
<td>-11.68</td>
</tr>
<tr>
<td></td>
<td>(5.00)</td>
<td>(8.49)</td>
</tr>
<tr>
<td>Pref-Vote Share:</td>
<td>-0.35***</td>
<td>-0.32***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.064)</td>
</tr>
<tr>
<td>Experience:</td>
<td>0.12**</td>
<td>0.24**</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>$R^2$:</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>Observations:</td>
<td>649</td>
<td>649</td>
</tr>
<tr>
<td>Groups:</td>
<td>387</td>
<td>387</td>
</tr>
<tr>
<td>F-statistic:</td>
<td>98.94***</td>
<td>8.62***</td>
</tr>
<tr>
<td>$\rho$:</td>
<td>0.42</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Notes: ** represents $0.01 \leq p < 0.05$, *** represents $p < 0.01$.
Heteroscedasticity-adjusted robust standard errors are in parentheses. All figures rounded to 2 decimal places.

to a drop of 0.12 places. Less than 10 years are needed for an individual to drop by 1 place on the party list, on average. Even though this is not in line with Hypothesis 2, it is not as counterintuitive as one might initially think (see the discussion below).

Having eliminated the only time-invariant variable in our model, we can now also estimate a fixed-effects model. The benefit of the fixed-effects model is that it automatically accounts for all time-invariant variables shielding the model from omitted variable bias - at least to an extent: provided that the omitted variables are time-invariant (Asteriou, Hall, 2011, p. 419). We thus re-estimate the model using fixed effects. Stata has an inbuilt fixed-effects functionality which means that we do not have to create dummies for each individual and include them in the model. The two models are compared in Table 3 (full output in the Appendix, Figure 10).

Before we look at the coefficients of the fixed-effects model, we run the Hausman test to determine whether to use fixed or random effects (Asteriou, Hall, 2011, pp. 420-421; Baltagi, 1995, pp. 68-73). To this end, we need to re-estimate the models again because Hausman test does not support robust standard errors. We thus run the two models again with ordinary standard errors. The Hausman test
(see the Appendix, Figure 11) rejects the null hypothesis that the preferred model is the random effects one at the highest, 1% level of significance. We are thus advised by the test to use the fixed effects model. Before interpreting the coefficients still, we have performed a test to determine whether time fixed effects are needed in addition to the individual fixed effects (Torres-Reyna, 2011). The test failed to reject the null hypothesis that all the coefficients for the time dummies are equal to 0 (p=0.45) and we thus conclude that we do not need time fixed effects (Appendix, Figure 12). Looking at the coefficients of the fixed effects model then, we note that the party-line voting coefficient is no longer distinguishable from 0. This means that we no longer find support for Hypothesis 1 and are left with sound support for Hypothesis 7. This is further analysed in the Discussion subsection which follows below.

9.3 Assumptions testing

Having settled on the final model then, we now turn to assumption testing. Panel-data linear regression models rest on the same set of assumptions as simple OLS models: no outliers or influential data, normality of error terms, homoscedasticity, no multicollinearity, linear relationship between the regressors and the response variable and no serial correlation. We scrutinise the first 5 assumptions but do not worry about the last one. Torres-Reyna (2011) claims that serial dependence only poses a problem with macro panels with a long time series of 20-30 years. We, however, only have 5 time points. This means that our model is unlikely to suffer from serial correlation. We nevertheless do test for autocorrelation using the XTSERIAL command and fail to reject the null of hypothesis of no autocorrelation. The test, however, might be suffering from low power given the short time series (see Figure 17 in the Appendix). Finally, we assume that the model is correctly specified, that the variables are accurately measured and that no variables are omitted. As already argued, this is something we must assume. We do know that some variables are most likely omitted from our model but we do believe that the final model we settle on is the best we can get given the data we have at hand.

9.3.1 Influential observations and outliers

We have estimated the fixed effects model using the reg option in Stata and dummies for each individuals. This way, we could estimate the DFITS option which allows one to identify outliers and influential observations. Using the standard cutoff point of 2 and -2, we have excluded 160 observations and re-estimated the models again. The new models were not different from the full-sample models as far as the signs of the coefficients and their significance is concerned. The only difference was that the party-line voting in the restricted random-effects model was no longer significant at 5% but 10% level of significance. Most of the assumptions remained violated (see below) even after the restricted-sample models were estimated. This is why, we have decided to keep all the observations in our sample and proceed with the analysis of the assumptions.
9.3.2 Normality of the error terms

One of the crucial assumptions of linear regression is that the error terms are normally distributed. We check for this assumption by looking at the residuals. The residuals were obtained using the predict, e command in Stata. Figure 3 shows the resulting QQ-plot.

Provided that the residuals are normally distributed, the QQ-plot should follow a reasonably straight line. Curvilinear patterns represent departures from normality. As Figure 3 clearly indicates, our residuals do follow quite a clear curvilinear pattern which means that our residuals are most probably not normally distributed. This might be because the true data generating process is not linear in its character or for various other reasons. Normality of the residuals might sometimes be improved when some of the variables included in the model are transformed. And in our model, we might have a reason for suspecting that at least one variable needs to be transformed. Experience is often thought to have a diminishing-return-like relationship with income, productivity, promotion, etc. What this means is that while experience does indeed matter initially, improving one’s chances to succeed, its effect begins to flatten out the more experienced one becomes. We do attempt to transform some of our variable but only after we have inspected the partial regression plots (see the Linearity subsection below). This transformation, however, does not improve the normality plot of the residuals.

Finally, we perform the Jarque Bera normality test and reject the null hypothesis of normality at the highest 1% level of significance ($p < 0.0001$). We should
note that in large samples, any small deviation from normality will likely become significant. In our case, however, normality is most likely truly violated, as the QQ plot clearly indicates. The Linearity subsection below described how we handled the violation of the normality of residuals assumption.

9.3.3 Homoscedasticity

Further important assumption is that the error terms have a constant variance. We check whether this assumption holds by looking at the scatterplot of the fitted values versus the residuals. The consequence of having this assumption violated is that the point estimates are still unbiased and consistent, but no longer efficient (Asteriou, Hall, 2011, p. 113). Moreover, the standard errors might either be underestimated or inflated, making t- and F-statistics unreliable, limiting our the possibilities to do meaningful inference.

What we should ideally see in the graph is no clear patterns - that is, equal spread of residuals across the different fitted values. What we do not want to see is a funnel-like pattern which would indicate that heteroscedasticity is present. Figure 4 shows the scatterplot of the fitted values and residuals.

Figure 4 shows a clear, funnel-like shape, which is a strong indicator that the homoscedasticity assumption has been violated. To further investigate whether this is indeed the case, we run a modified Wald test for heteroskedasticity (xttest3 in Stata). At the 1% level of significance, we reject the null hypothesis of homoscedasticity and conclude that heteroscedasticity is indeed present (Appendix,
To remedy, for the violation of this particular assumption, one can estimate heteroscedasticity-adjusted robust standard errors rather than ordinary standard errors. Robust standard errors tend to inflate (or deflate) the estimates of standard errors slightly to adjust for heteroscedasticity. We have been applying robust standard errors from the very start suspecting that heteroscedasticity might pose a problem to our models. There is thus no need to estimate a new model.

### 9.3.4 Multicollinearity

As already indicated, we have tried to eliminate the variables that often cause multicollinearity at an earlier stage. It is not possible to estimate the variance inflation factor after running an `xtreg` in Stata. One can estimate a fixed-effects regression the conventional way - using dummies for every individual. That however, makes the computation of the VIF challenging for the software. We thus estimated a `reg` model, using all the explanatory variables but leaving out the individual dummies and calculated the VIF. With none of the VIF values exceeding 1.5, we can safely conclude that multicollinearity is not present in our model (Appendix, Figure 14).

### 9.3.5 Linearity

In order to investigate whether the relationship between our explanatory variables and the dependent variable of interest is linear in its character, we look at the partial regression plots (Stata, 2013). Figure 5 shows the partial regression plots for the initial model.

As Figure 5 clearly indicates, the relationship between preferential vote share and party-list rank at t+1 does not seem to be linear. Instead, the relationship could be thought of as a negative square root of the vote share variable (L-shape). Those who receive few preferential votes tend to receive all kinds of party-list ranks, but those who receive a lot of preferential votes tend to only receive good positions. Furthermore, the partial regression plot for experience seems to resemble a U-shape. This is intuitive again, those with little experience tend to receive worse party-list ranks and their ranks improve as their experience increases until a certain threshold is reached. After this threshold is passed, the aspirants begin to receive worse ranks again (U-shape). In order to model this, we take a square of Experience. The newly estimated model yields much better looking partial regression plots (Figure 6).

Furthermore, the new model also has higher $R^2$ than the initial fixed effects model (we are hesitant to use $R^2$ statistics as a mode of comparison as these are tied to the degree of variability of the response variable). What has also changed, however, is that the party-line voting coefficient is statistically significant again. The two models are compared in Table 4. We prefer the second model because
it suits the data at hand better and it is also intuitive. The coefficients for the experience and vote-share variables get more difficult to interpret, but the model gains by achieving a better fit. After having estimated the new model, we have checked all the aforementioned assumptions again, but there has been only a very
Table 4: Initial Fixed-effects Vs. transformed Fixed-effects Models, Party List
Rank at t+1 is the dependent variable

<table>
<thead>
<tr>
<th>Model:</th>
<th>FE2</th>
<th>FE1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept:</td>
<td>21.5***</td>
<td>15.39**</td>
</tr>
<tr>
<td></td>
<td>(7.67)</td>
<td>(7.89)</td>
</tr>
<tr>
<td>Party-Line Voting:</td>
<td>-16.27**</td>
<td>-11.68</td>
</tr>
<tr>
<td></td>
<td>(8.07)</td>
<td>(8.49)</td>
</tr>
<tr>
<td>Pref-Vote Share:</td>
<td>155***</td>
<td>-0.32***</td>
</tr>
<tr>
<td>(-Square Root of Pref-Votes)</td>
<td>(0.35)</td>
<td>(0.064)</td>
</tr>
<tr>
<td>Experience:</td>
<td>0.02***</td>
<td>0.24**</td>
</tr>
<tr>
<td>(Experience Squared)</td>
<td>(0.006)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>R²:</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>Observationthesiss:</td>
<td>649</td>
<td>649</td>
</tr>
<tr>
<td>Groups:</td>
<td>387</td>
<td>387</td>
</tr>
<tr>
<td>F-statistic:</td>
<td>7.16***</td>
<td>8.62***</td>
</tr>
<tr>
<td>ρ:</td>
<td>0.60</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Notes: ** represents 0.01 ≤ p < 0.05, *** represents p < 0.01.

Heteroscedasticity-adjusted robust standard errors are in parentheses. All figures rounded to 2 decimal places. FE2 model contains 2 transformed variables. These are specified in brackets underneath the original variables. The full output of both models can be found in the Appendix.

A marginal improvement as far as normality of the residuals and homoscedasticity are concerned. We thus still rely on homoscedasticity-adjusted robust standard errors. To remedy for the violation of the normality of residuals assumption, we rely on bootstrapped standard errors. Bootstrapping does not rely on the normality assumption in order to calculate standard errors, t-values and F-values. For the comparison of coefficients, see Table 5.

The two models presented in Table 5 are thus to two models we finally settle on. Both models still suffer from violation of some of the crucial assumptions, but we have tried to take care of those by applying appropriate remedies. Note that this does not mean that we no longer need to be worried about them: we must be very cautious, when making inferences onto a wider population. The final model, however, is the best we can get.

9.4 Alternative model specifications

Finally, we have also attempted to estimate a number of complementary models (not reported, available from the author). First, we have divided the sample into two - one containing left-leaning parties and the other containing centre-right
### Table 5: Transformed Fixed-effects (Robust standard errors) Vs. transformed Fixed-effects (bootstrapped standard errors) Models, Party List Rank at t+1 is the dependent variable

<table>
<thead>
<tr>
<th>Model:</th>
<th>FE (Robust SE)</th>
<th>FE (Bootstrapped SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept:</td>
<td>21.5*** (7.67)</td>
<td>21.5*** (6.96)</td>
</tr>
<tr>
<td>Party-Line Voting:</td>
<td>-16.27** (8.07)</td>
<td>-16.27** (7.34)</td>
</tr>
<tr>
<td>-Square Root of Pref-Votes:</td>
<td>1.55*** (0.35)</td>
<td>1.55*** (0.3)</td>
</tr>
<tr>
<td>Square of Experience:</td>
<td>0.02*** (0.006)</td>
<td>0.02*** (0.006)</td>
</tr>
<tr>
<td>( R^2 ):</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>Observations:</td>
<td>649</td>
<td>649</td>
</tr>
<tr>
<td>Groups:</td>
<td>387</td>
<td>387</td>
</tr>
<tr>
<td>F-statistic:</td>
<td>7.16***</td>
<td>27.07***</td>
</tr>
<tr>
<td>( \rho ):</td>
<td>0.60</td>
<td>0.60</td>
</tr>
</tbody>
</table>

**Notes:** ** represents \( 0.01 \leq p < 0.05 \), *** represents \( p < 0.01 \). Standard errors are in parentheses. All figures rounded to 2 decimal places.

Parties. We have done so in order to investigate whether some of the variables that matter to re-selection on average also matter for each spectrum individually. Most of the non-significant coefficients remained non-significant in the two left-right models. One’s biological sex, however, turned out significant in the left-leaning model. Being a lefty female legislator is associated with receiving a lower party-list spot at t+1, on average, than being a man, controlling for all other variables in the model. This finding is indeed interesting because it provides some evidence for the frequent argument that party recruiters discriminate against women just because of their gender, seeing them less predictable than men. It is surprising, however, that it is the left-leaning parties rather than their right-wing counterparts that engage in this kind of discrimination. Left-wing parties are generally perceived as being more progressive as far as women’s representation is concerned. This, however, often does not hold in post-Communist societies where left-wing parties, often remnants of the Communist parties, tend to be much more Conservative than their Western-European counterparts. The sex variable remains insignificant in the right-wing model.

Party-line voting, moreover, seems to matter much more in left-leaning parties. Even though the coefficient is significant in both models, the magnitude at which
party-line voting improves one's position on the party list in left-leaning parties is much greater than in their right-wing counterparts. This also shows that the left-leaning parties in the Czech Republic are much more centralised and place much more emphasis on party loyalty.

Finally, when included in the models, the education dummies turn out non-significant. This further validates our claim that different factors are likely to determine one's success in different political selections and justifies our framework for a more universal theoretical framework on political selections that is sensitive to this fact. Education is often spelled out as one of the most crucial determinants (together with income and occupation) of political recruitment and initial political selection. Yet, it does not seem to matter much at the re-selection stage. As theoretically argued, this might be because party recruiters have much better proxies at their discretion to evaluate each aspirant.

X. Discussion

In this study, we were set to investigate the factors that determine which party-list rank aspirants who seek re-election receive. By focusing on party-list rank as a proxy for how successful each aspirant has been at the re-selection stage, we have attempted to overcome the fundamental problem of selection research that the pool of unsuccessful candidates is usually unobserved. In case of re-selection (selection of those MPs who have been elected in the previous term), we do know all those who do not re-appear on the party and do have all the relevant data on them. What we do not know, however, is whether they have entered the aspirants’ pool and were unsuccessful or whether they have quit Politics voluntarily. This is why, we believe that focusing on party-list rank can tell us a lot about how each individual fared at the re-selection process (assuming that all those who enter the aspirants’ pool do desire to be placed as high as possible).

Having this puzzle in mind, we have made use of the principal-agent framework explaining political recruitment and adapted it to our instance of political selection. This framework places a lot of emphasis on party recruiters who rely on proxies in order to decide which of the aspirants will better further their party’s ends. We have hypothesised that one’s party-line voting, age, experience, gender, preferential votes at the previous elections, party membership and whether one has held a senior position during the last term could be good determinants of which party rank each aspirant will receive. Most of these variables (party-line voting, party membership, senior position) are mere proxies for trust. By entrusting a senior position to somebody, party gatekeepers invest certain degree of trust into a given individual. At the end of the term (or even during the term), they likely evaluate whether this trust has paid off or was a mere miscalculation. The senior position variable will capture the initial trust, but will have very little to say about the post-term analysis and deliberation. These variables are thus naturally correlated with an unmeasurable variable trust, which is omitted from
the model. When omitted, trust (and, possibly, other omitted variables, if any) becomes part of the error term which in turn becomes correlated with the regressors (violating one of the assumptions of the random effects model). The model thus suffers from methodological weaknesses from the very start. This is due to the very complex nature of the social reality we attempt to model.

Turning to the empirical results, we have anticipated that voting along the party line will improve one's position on the party list. And indeed - we have found sound empirical evidence that this is the case. The coefficient turned out non-significant in one of the models we have estimated but this has changed after we have transformed some of the other variables which suffered from lack of linearity. The relationship between party-line voting and party list rank is negative, which means that the more similar one's voting with the party average, the better (higher) rank one receives. This is not counterintuitive and provides some support for the agent-principal framework. It truly seems to be the case that the recruiters do assess one's party loyalty when choosing who is to get a better spot on the party list and who is not.

The second prediction was the relationship between experience (defined as the number of years spent in the legislature) and party-list rank. As already noted, the association is positive which goes against the initial prediction. This, however, is not counterintuitive. The longer one stays in the legislature, the more threat this person might pose to the party leadership. Moreover, party recruiters might feel that the old cadres should give opportunity to new faces in the process of natural party turnover. Furthermore, party recruiters might feel confident about placing older MPs onto lower places because their name recognition might anyway help them to attract preferential voting and move them up the list. Finally then, the more experienced politicians might demand to be placed into non-electable seats when desiring to leave Politics (check out Figure 5). Parties might want to keep their name on the party lists for the sake of name recognition one last time. These few individuals might then drive the average to positive values, especially if they are given very low ranks. The relationship between experience and party-list rank remained significant in all the models. After inspecting the partial regression plots, however, it became apparent that the relationship between experience and party-list rank might better be modelled as a U-shaped rather than a linear one. We have thus taken a square of experience, which yielded a better partial regression plot. The coefficient remained significant.

We have also hypothesised that the share of preferential votes might also be an important factor that party recruiters must consider. This is because preferential voting works as some kind of an accountability mechanism. If voters like a particular representative, they award him/her with their preferential votes. This, in the eyes of the party, legitimises the actions of the individual in question - even if they might be at odds with party interests at times. As thought, preferential vote share is a strong predictor of party-list rank at t+1. The more votes one receives, the better rank one gets before the next elections. After inspecting the
partial regression plots, we have realised that the relationship is probably not
linear. This is because there are many people who’s share of the preferential votes
is small. These people tend to receive all kinds of party-list ranks and thus we
have a pile of observations above the low values of preferential voting. As the
share of preferential votes increases we only get low values on the dependent
variable (they receive better spots). We thought that the relationship could thus
be better captured as a negative square root of preferential voting. And indeed,
this model specification yielded a better model fit while the coefficient remained
significant.

The coefficients for the remaining variable we have thought might have an ef-
fect on one’s party-list rank were not statistically distinguishable from 0. We thus
do not have enough evidence to either reject or confirm them. This includes the
effect of biological sex the coefficient of which was persistently non-significant
(except for left-leaning parties). Same applies to variables representing position
of responsibility (Committee chair, speaker and vice-speaker).

The reader should note, however, that the final model we have estimated vio-
lates some of the assumptions it is built upon. Most notably, the assumption of
normally distributed residuals is violated and data transformation has not helped
to provide a remedy for this. The violation of this assumption does not bias the
point estimates but induces bias to standard errors, t-value and F-values and
confidence intervals. We have tried to remedy for this by using bootstrapped
standard errors which do not rely on the normality assumption. This, however,
is only a weak remedy. Finally, the model suffers from heteroscedasticity which
also has an adverse effect on the standard errors (and thus t-values and con-
fidence intervals). To remedy for this, we have consistently made use of robust
standard errors.

XI. Conclusion

In this Bachelor’s thesis, we have employed panel data regression analysis, to
shed some light on the under-researched issue of political selection. We have
focused on re-selection onto party lists of those who have been elected in the
previous term. We have applied the principal-agent theoretical framework, often
employed in the political-selection literature, to determine which factors might
matter to one’s re-selection. Using panel data on more than 300 Czech legisla-
tors, we have found evidence that party-line voting is a strong predictor of one’s
party-list rank at t+1. The more one’s voting record resembles that of the average
party vote, the better spot one receives on the party list at t+1. One’s share of
the total party preferential vote is also positively tied to one’s rank on the party
list at t+1. We have argued that this is because preferential voting works as
some kind of an accountability mechanism. The higher one’s share, the strongest
the signals for the party that one should be promoted and given more spotlight.
Finally, one’s experience defined as the number of years one has worked as a
legislator is negatively associated with one’s rank on the party list at t+1.

This thesis thus contributes to the emerging literature on political selection. Methodologically, it tries to overcome what is often dubbed the fundamental problem of selection research of not being able to observe the pool of unsuccessful aspirants. By treating party-list rank as a proxy for how well one fared at the re-selection process, we believe we have shed some light on the dynamics of the re-selection process. Even though we have not been able to confirm/reject all our hypotheses, we believe that we have set the stage for further testing on larger samples.

XII. Bibliography

12.1 Theoretical Framework


Folke, Olle, Rickne, Johanna, "Female representation but male rule? Party competition and the political leaking pipe", IDEA Stockholm, Online: 2012.


Rahat, Gideon, 'Which Candidate Selection Method is the Most Democratic', *Government and Opposition*, 44.1, 2009: 68-90.


### 12.2 Methodological Design


Forsberg, Lars, Lecture slides in Econometrics B, HT2013, Uppsala University.

Stata: Data Analysis and Statistical Software: Various online resources, http://www.stata.com


### 12.3 Other Online Resources

### XIII. Appendix

Figure 7: Full Random-effects Model, Party-List Rank at t+1 is the dependent variable

| Variable | Coef. | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|----------|-------|-----------|-----|-----|---------------------|
| Voting_1 | -12.00131 | 4.892538 | -2.45 | 0.014 | -21.50051 to -2.412112 |
| Pref_Share | -0.3562945 | 0.0366577 | -9.77 | 0.000 | -1.4301423 to -0.2804467 |
| Age | 0.0534684 | 0.0285563 | 2.60 | 0.009 | 0.0131788 to 0.093758 |
| CommChair | -0.4901794 | 0.5658966 | -0.88 | 0.379 | -1.6671766 to 0.685176 |
| SpeakerVS | -0.238388 | 0.194595 | -0.57 | 0.570 | -1.060496 to 0.5837198 |
| Experience | -0.1907022 | 0.504368 | 1.37 | 0.095 | -0.153836 to 0.1973921 |
| Sex | 0.4858086 | 0.5164984 | 0.94 | 0.347 | -0.5265097 to 1.498127 |
| _cons | 14.04783 | 4.528504 | 3.10 | 0.002 | 5.172122 to 22.92353 |

| | | | | |
| sigma_u | 2.5639714 |
| sigma_e | 3.8725972 |
| rho | 0.418042904 (fraction of variance due to u i) |
Figure 8: Full Random-effects Model: Breusch and Pagan Lagrangian multiplier test for random effects

\[ \text{Breusch and Pagan Lagrangian multiplier test for random effects} \]

\[ \text{Party\_List\_Rank\_t1[ID,t] = Xb + u[ID] + e[ID,t]} \]

Estimated results:

<table>
<thead>
<tr>
<th></th>
<th>Var</th>
<th>sd = sqrt(Var)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Party_L_1</td>
<td>17.36624</td>
<td>4.167282</td>
</tr>
<tr>
<td>e</td>
<td>9.440854</td>
<td>3.072597</td>
</tr>
<tr>
<td>u</td>
<td>6.573949</td>
<td>2.563971</td>
</tr>
</tbody>
</table>

Test: \( \text{Var(u) = 0} \)

\[ \text{chi}^2(01) = 17.12 \]

\[ \text{Prob > chi}^2 = 0.0000 \]

Figure 9: Restricted Random-effects Model, Party-List Rank at \( t+1 \) is the dependent variable

\[ \text{. xtreg Party\_List\_Rank\_t1 Voting\_l Pref\_Share Experience, re ve(robust)} \]

Random-effects GLS regression

Number of obs = 649
Number of groups = 387

R-sq: within = 0.0593
between = 0.1171
overall = 0.1103

Obs per group: min = 1; avg = 1.7; max = 5

corr(u_i, X) = 0 (assumed)

Wald chi2(3) = 98.94
Prob > chi2 = 0.0000

(Std. Err. adjusted for 387 clusters in ID)

| Party\_List\_l | Robust Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|----------------|--------------|-----------|---|-----|----------------------|
| Voting\_l      | -10.05755    | 4.995716  | -2.01 | 0.044 | -19.94898 | -10.166276 |
| Pref\_Share    | -3.354399    | 0.4362958 | -7.66 | 0.000 | -4.215784 | -2.493014 |
| Experience     | 0.1185514    | 0.0536911 | 2.11  | 0.037 | 0.013187  | 0.223784  |
| _cons          | 14.7006      | 4.65651   | 3.16  | 0.002 | 5.574089  | 23.82719  |

\[ \text{sigma_u = 2.5904501} \]

\[ \text{sigma_e = 3.0653299} \]

\[ \text{rho = 0.41662566 \ (fraction of variance due to u_i)} \]
Figure 10: Restricted Fixed-effects Model, Party-List Rank at t+1 is the dependent variable

```
.xtreg Party_List_Rank_t1 Voting_1 Pref_Share Experience, fe vce(robust)
```

<table>
<thead>
<tr>
<th>Fixed-effects (within) regression</th>
<th>Number of obs</th>
<th>= 649</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: ID</td>
<td>Number of groups</td>
<td>= 387</td>
</tr>
<tr>
<td>R-sq: within = 0.0884</td>
<td>Obs per group: min = 1</td>
<td></td>
</tr>
<tr>
<td>between = 0.0795</td>
<td>avg = 1.7</td>
<td></td>
</tr>
<tr>
<td>overall = 0.0767</td>
<td>max = 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F(3,386) = 8.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prob &gt; F = 0.0000</td>
<td></td>
</tr>
</tbody>
</table>

> corr(u_i, Xb) = -0.8796

(Std. Err. adjusted for 387 clusters in ID)

| Party_List-Rank | Robust Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|-----------------|--------------|-----------|---|------|----------------------|
| Voting_1        | -11.6777     |  8.467603 | -1.38 | 0.170 | -28.3655 | 5.009947 |
| Pref_Share      |  -0.3172367  |  0.0642902 | -4.93 | 0.000 | -0.4438394 | -0.1908339 |
| Experience      |  0.2428357   |  0.0768868 | 3.19  | 0.002 | 0.0932392 | 0.3924322 |
| Cons            |   15.38822    |  7.889866 | 1.95  | 0.052 | -1242727 | 30.90071 |
| sigma_u         |   3.8529714   |       |
| sigma_e         |   3.0653299   |       |
| rho             |   0.61239219  | (fraction of variance due to u_i) |

Figure 11: Hausman Test

```
.hausman fe re
```

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
<th>Differe ce</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voting_1</td>
<td>-11.6777</td>
<td>-10.05755</td>
<td>-1.620223</td>
<td>7.176046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pref_Share</td>
<td>-0.3172367</td>
<td>-0.3584399</td>
<td>0.0332032</td>
<td>0.076695</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>0.2428357</td>
<td>0.1185514</td>
<td>0.1242843</td>
<td>0.0499233</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

```
chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 25.57
Prob>chi2 = 0.0000
```

43
Figure 12: Time Fixed-effects Test

```stata
. testparm i.Election
```

( 1) 1998.Election = 0
( 2) 2002.Election = 0
( 3) 2006.Election = 0
( 4) 2010.Election = 0

```
F( 4, 255) = 0.93
Prob > F = 0.4488
```

Figure 13: Heteroscedasticity Test

```stata
. xtest3
```
 Modified Wald test for groupwise heteroskedasticity in fixed effect regression model

```
H0: sigma(i)^2 = sigma^2 for all i
```

```
chi2 (387) = 8.9e+30
Prob>chi2 = 0.0000
```

Figure 14: Variance Inflation Factor

```stata
. estat vif
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF</th>
<th>1/VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>1.19</td>
<td>0.843320</td>
</tr>
<tr>
<td>Voting_1</td>
<td>1.12</td>
<td>0.892660</td>
</tr>
<tr>
<td>Pref_Share</td>
<td>1.09</td>
<td>0.913285</td>
</tr>
<tr>
<td>Mean VIF</td>
<td>1.13</td>
<td></td>
</tr>
</tbody>
</table>
Figure 15: Restricted Fixed-effects model with two transformed regressors + robust SEs

<table>
<thead>
<tr>
<th>Fixed-effects (within) regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: ID</td>
</tr>
<tr>
<td>R-sq: within = 0.0819</td>
</tr>
<tr>
<td>between = 0.1282</td>
</tr>
<tr>
<td>overall = 0.1179</td>
</tr>
<tr>
<td>Obs per group: min = 1</td>
</tr>
<tr>
<td>avg = 1.7</td>
</tr>
<tr>
<td>max = 5</td>
</tr>
<tr>
<td>F(3, 386) = 7.16</td>
</tr>
<tr>
<td>Prob &gt; F = 0.0001</td>
</tr>
<tr>
<td>corr(u_i, Xb) = -0.0016</td>
</tr>
</tbody>
</table>

(Std. Err. adjusted for 387 clusters in ID)

|     | Robust   | Coef.  | Std. Err. | t   | P>|t| | [95% Conf. Interval] |
|-----|----------|--------|-----------|-----|-----|---------------------|
| L2  |          |        |           |     |     |                     |
| Voting | -16.27221 | 8.070163 | -2.02 | 0.044 | -32.13819 | -40.52302 |
| MinPref_Votes | 1.552718 | 0.3527564 | 4.40 | 0.000 | .8591534 | 2.246282 |
| Exp2  | .0163234 | .0056494 | 2.89 | 0.004 | .005216 | .0274307 |
| cons  | 21.50474 | 7.67484 | 2.80 | 0.005 | 6.415021 | 36.59447 |
| sigma_u | 3.7450689 |
| sigma_e | 3.0762338 |
| rho   | .59721256 | (fraction of variance due to u_i) |
Figure 16: Restricted Fixed-effects model with two transformed regressors + bootstrapped SEs

<table>
<thead>
<tr>
<th>Fixed-effects (within) regression</th>
<th>Number of obs</th>
<th>649</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: ID</td>
<td>Number of groups</td>
<td>387</td>
</tr>
</tbody>
</table>

R-sq: within = 0.0819
between = 0.1282
overall = 0.1179

Observe per group: min = 1
avg = 1.7
max = 5

corr(u_i, x_i) = -0.0016

Wald chi2(3) = 27.87
Prob > chi2 = 0.0000

(Replications based on 387 clusters in ID)

| L2  | Observed Coef. | Bootstrap Std. Err. | z   | P>|z|       | Normal-based [95% Conf. Interval] |
|-----|----------------|---------------------|-----|----------|-----------------------------------|
| Voting_1 | -16.27221 | 7.342611 | -2.22 | 0.027 | -30.66346 -1.880956 |
| MinPref_Votes | 1.552710 | 0.833706 | 5.12 | 0.000 | 0.9501225 2.147313 |
| Exp2 | 0.0183234 | 0.0059817 | 2.77 | 0.006 | 0.0047563 0.0328904 |
| _cons | 21.58474 | 6.964947 | 3.09 | 0.002 | 7.853699 35.15579 |

sigma_u = 3.7450689
sigma_e = 3.0762638
rho = 0.59711256 (fraction of variance due to u_i)

Figure 17: Autocorrelation Test

```
.xtserial L2 Voting_1 Experience Pref_Share
```

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
F( 1, 67) = 0.265
Prob > F = 0.6086