Erasmus School of Economics

“Douze points!”

A research paper on reducing the effect of the familiarity factors by type of voting system

Research Paper presented by:

Andre Bos
409580

Supervisor: **dr. G.D. Granic**
Second assessor: **P. van Bruggen**

Rotterdam, The Netherlands
July 2019

Abstract:

Many studies have been done about the Eurovision Song Contest. This study has investigated which external factors have an effect on the difference between the number of points a particular voting country assigns and the quality of the song. The paper also tries to find whether changing the voting system reduces the influence of these external factors on that difference.
After analyzing voting data from 1975-2018 it appeared that there are several external factors that affect the voting bias. This paper revealed a significant effect that changing the voting system reduces the influence of most of those external factors. The voting system where a jury of experts is been used, reduces the effect of most of the external factors on the difference.

Keywords:

Eurovision Song Contest, Expert jury, Familiarity factors, Voting bias, Voting systems

Acknowledgements:

I would like to thank my family and girlfriend for supporting me during my studies, Jeremy Daynes for revising my thesis on language, the professors who taught me everything I need to know about (behavioral) economics, and dr. G.D. Granic who supervised me and helped me with my thesis.

Table of content

1. Introduction

“Douze points!”/”Twelve points!” a statement that would like to be heard by every Eurovision Song Contest participant. This means that their performance receives the highest number of points that can be given by a particular jury. The Eurovision Song Contest itself is the biggest non-sporting event in the world, and was viewed by 182 million people all over the world in 2019 (Spiteri, 2019). In this competition each participating country submits a representing song, which is evaluated by the juries of other participating countries. The points given by countries, in 2019, were assigned by a national expert jury and by televoters/laypeople. Televoting means that residents of a country can vote for their favorite participant via telephone or other medium. The person with the highest total number of points in the grand finale may call himself the winner of the Eurovision Song Contest (ESC) until the next edition starts again. However, various studies show that the winner has not always put down the best performance or sang the best song in qualitative terms, but other factors also may play a role in the number of points a participant receives. For example, it was written in the Economist (2013) that "the contest has always been about more than music". According to the article, language has played a major role in winning the ESC. English songs have won 24 times and French songs 14 times, despite a rule determined by the organization, that participants had to sing in their own language in two periods (1956-1965 and 1978-1998). The question here is whether there is a connection between the number of points that participants receive and, for example, a factor such as “language of the song performed”.

In addition to this discovery, there are more external and internal factors that may affect the number of points a voting country assigns to a participating country, according to several studies. In other words, there are factors which are somewhat related to ESC and factors which are totally unrelated to it. Factors that are unrelated to the ESC could be, for example, whether countries are neighboring countries, have common religion or if there are minority groups of a participant country living in the voting country[1]. However, another question that needs answering is what would happen with the effect these external factors have on the number of points a voting country assigns to a participating country, if a different type of voting system would be used. The organisation of the ESC has changed the voting rules a couple of times. Countries had to use expert juries, laypeople juries and, as discussed, in anno 2019 they had to use a combination of both to decide which song they thought was best.

This paper will examine in particular if experts are less prone to be influenced by unrelated factors, or whether laypeople are less prone to be influenced by those factors. In other words, the effect of a change in type of voting system on the outcome of the song contest will be analyzed and whether this change means that external factors have less effect on subjectivity[2] of voters when they vote for a participant.

[1] These external factors, also known as familiarity factors in this paper, are expected to capture a common taste. The familiarity factors themselves will be discussed in chapters 3 and 4 and whether they can capture a common taste will be discussed in chapter 7.

[2] Also called voting bias in this paper, more about this in chapter 3 and 4.

The research question is therefore as follows:

To what extent is a change in voting system successful in reducing the effect of external factors on the outcome of the ESC?

It is this question that is central to this paper, because it is obvious that in competitions such as the ESC performances should be judged as objectively as possible. When a participant is singing a song, it should be about nothing but the quality of the song. For that reason, this paper will be testing specifically if and which external factors affects the level of subjectivity voting country A has when assigning points to participating country B. In addition there will be also tested which type of voting system has an effect on that subjectivity level and whether a change of voting system will reduce or increase the effect that external factors have on the level of subjectivity. To test these effects data of the ESC from 1975-2018 will be used in several econometric models.

The results of those models show that several external factors are having a significant effect on the level of subjectivity. Secondly, evidence is found that the type of voting system used by a voting country affects the level of subjectivity it has when assigning points to participating country B. When a country uses a jury of experts this will increase the level of subjectivity of that specific jury, compared to a jury of laypeople or a combination system. However, evidence also showed that changing the voting system to a jury of experts will significantly reduce the effect of most of external factors on the level of subjectivity, compared to the other two voting systems. So this means that experts are less prone to external factors like the one's included in the models.

These results thus reveal which external factors affect voting behaviour of laypeople and professionals juries, thereby making a contribution to literature about how people vote collectively and what influences certain voting behavior. Next to this, the results of this paper should also be interesting to the organization of the ESC. The research examined that a jury of experts is best to use, in order to minimize the effect of external factors. The results, in a broader scope, can also be of value to other election organizations, such as the parliamentary elections, because it shows that experts are most objective in recognizing quality. As such, it demonstrates that experts are least affected by external factors and therefore makes it more likely that they are able to identify who, objectively speaking, should win the elections.
Finally, this research will be an extension to existing literature on voting biases that play a role in the ESC, by using data from the moment a new voting system was introduced in 2009.

The rest of this paper is organized as follows. Chapter 2 explains the background and rules of the Eurovision Song Contest. Chapter 3 shows relevant papers from this research area which are focused on voting biases in the ESC or other comparable competitions. Chapter 4 and 5 describe the data and the methodology used in this research. Chapter 6 contains the models and descriptions of the results derived from the models. Finally, chapter 7 summarizes this research and gives suggestions for further research.

2. Background Eurovision Song Contest

ESC data will be used to answer the research question. The reason for this is that all data regarding the results have been made public in the past. In addition, juries from different countries can also have cultural, geographical and/or linguistic similarities, so they are easier to compare. This makes ESC data ideal for use in this study. Before looking at previous studies on this topic, this chapter will show information about rules and changes to those rules.

The ESC is an international song competition that started in 1955. At the time, only 7 countries participated, whereas it has now grown to more than 40 countries (Eurovision, 2016). The similarity between these countries is that the participants have to be members of the European Broadcasting Union (EBU). There are now 56 countries represented in the EBU, both from the European continent and from outside the continent (EBU, 2019). Every year the EBU invites all members to participate in the Eurovision Song Contest. In addition, as a "special guest", Australia has been invited since 2015 to participate.
Because so many countries are participating now, it was decided in 2004 to introduce a semi-final (and since 2008 two semi-finals), so that the number of participants in the final is limited to 26 participants (Budzinski & Pannicke, 2017). There are a couple of countries that already have a guaranteed ticket in the final, and therefore do not have to participate in the semi-finals. Originally these were five countries (France, Spain, Germany, UK and the hosting country) and in 2011 this was expanded to six when Italy was added. The rest of the countries have to earn their ticket by, since the introduction of two semi-finals, being among the 10 best performances in one of those two semi-finals (Budzinski & Pannicke, 2017). The hosting country has already a final ticket, because it is the winner of last year. In other words, the winner of a specific edition of the ESC is allowed to organize the next edition and thus already receives a ticket for the final of that edition.
In addition to changes in rules about the number of participants, the point system has also changed over the years. The point system, as it is still used today, was introduced in 1975. This points system runs from 0 to 12 points, with the participant receiving the highest total number of points of all countries as the winner (Eurovision, 2019). A country gives 12 points to a country they like best, 10 points to number two, 8 points to number three, and 7 to 1 point to numbers 4 to 10. All participants ranked lower than rank 10, do not receive points from that specific country (Budzinski & Pannicke, 2017).
As stated, the way points are obtained has changed a number of times. From 1975 to 1997 points were given by a national jury. The jury members individually awarded points to acts which they thought were the best. These points were aggregated which provided a ranking of songs for that particular national jury. This ranking then showed the number of points that a country would give to a participant (Haan, Dijkstra, & Dijkstra, 2005). In 1997 some countries experimented to use televoting instead of a jury. From 1998, every participating country used televoting[3].

[3] Some countries still used a professional jury, for example due to a low telecommunication coverage ratio or a too small population living in a country (e.g. San Marino) (O'Connor, 2010).

From 2009 points were divided by different combinations of televoting and a professional jury (Eurovision, Moscow 2009, 2019). In the first instance this was a score that was obtained by combining 50% of a televoting judgement and 50% of an expert jury judgment, but the average did not show how many points the jury had given and how many points the televoters had given. From 2016, this was distinguished from each other and participants received both televoting and professional jury points (Eurovision Voting, 2019).
If after the distribution of points no winner could be designated, because of a draw, for example, the number of times a participant had 12 points from other countries was decisive. If this was also the same, then the number of 10-points counted as decisive etc. Since 1975 this has only happened once when Sweden and France had an equal number of points and Sweden was ultimately declared the winner (Eurovision, Rome 1991, 2019). Since the introduction of a mixed system (televoting + professional jury), the number of points given by televoters is decisive.

Besides the rule changes about voting systems, there have also been other changes. Because this paper uses data from 1975 onwards, only the adjustments since 1975 that are most important for this study will be further explained.
Since the introduction of semi-finals, a participating country is still allowed to vote in the final, even when they didn't reach this themselves . If a country does participate in the final, the conditions for voting are that you cannot vote for your own participant as a country and that the professional jury must consist of 5 music professionals of their own nationality (Eurovision, Voting, 2019).
Another important rule that existed between 1978 and 1998 was that participants were not allowed to sing in a language other than the language spoken in their own country. Since 1999, this restriction no longer applies and countries that do not speak English, for example, may sing again in English (Spierdijk & Vellekoop, 2009).
If the song is performed by a group then this group may not be larger than 6 people (Spierdijk and Vellekoop, 2009).
Until 2013, the order of appearance was determined randomly. From 2013 this is in the hands of the executive producers of the hosting country. They determine, with the permission of the EBU, who is first and who is allowed to perform last, for both semi-finals and finals. It is still randomly determined who can perform in the first half and who can perform in the second half of the evening (Budzinski and Pannicke, 2017).
Finally, the organization of the ESC prescribes that the event must be "a non-political event". This means that the host country must ensure that political messages are not shown and that they are not referred to in the performances themselves (Eurovision Rules, 2019).

3. Literature review

This chapter provides an overview of previous studies that have been done on the effect of external factors and the effect of the type of voting system on the voting bias. The voting bias measures the deviation of the actual number of points given by country *i* to participants song *j* from the expected number of points based on quality. The precise formula to calculate 'quality' is given later on in chapter 4.

One of the first studies into the effect of external factors on the voting bias is by Yair (1995). He found that there were three blocks of countries that gave each other an above-average number of points. He analyzed data from 1975-1992 and from that came a Western Bloc, a Northern Bloc and a Mediterranean Bloc. According to Yair (1995), the Western Bloc exists because they have common historical and political factors, the Northern Bloc because of common cultural and linguistic factors, and the Mediterranean Bloc because of cultural factors. These '*familiarity factors'* were later examined again by Gatherer (2006). The extensive data set from 1975-2002 showed that there were another two voting blocs. Scandinavia and the Baltic states formed a Viking Empire Bloc. In addition, it was also apparent that Russia, Romania and former Yugoslavia countries were many times in agreement with each other (Warsaw Pact Bloc). These blocks arose in particular from the mid-1990s.

Clerides & Stengos (2006), and after them also Spierdijk and Vellekoop (2009), looked into the familiarity factors themselves and if they have an effect on voting behavior. In their research about the effect of geographical variables on the number of points that a participant receives, they controlled for factors such as culture, language, religion and ethnicity. These studies proved that there was a voting bias based on geographical characteristics, which is also in line with Yair (1995) and Gatherer (2004).

Fenn et al. (2006) also investigated voting patterns in the ESC. Data from 1992-2003 again showed that there were voting blocks. In addition, even smaller "collaborations" were found that were not necessarily in line with previously discovered voting blocks from Yair (1995) and Gatherer (2004). Based on their investigations, it would be expected that Turkey and Cyprus would vote for each other, given their common geographical and cultural characteristics. However, research by Fenn et al. (2006) showed that this was not the case. Countries that are geographically close to each other or that have the same cultural characteristics do not necessarily choose to vote for each other. The possible reason for this is because there could be more factors that affect the voting bias, which have not been taken into account so far.

One of those factors that play a role in voting behavior is the number of immigrants in a country. Dekker (2007) found that immigrants, for example, from Turkey living in Belgium, the Netherlands, France, Germany and Austria had an influence on the number of points that Turkey received at the ESC.

Another factor that has not been included in previous studies is the quality of the song. When this quality factor was included in the analysis, it appeared that this "quality factor" had the greatest effect on the results (Ginsburgh & Noury, 2008). Voting blocks could still be distinguished, but only on the basis of linguistic and cultural characteristics. Therefore, in contrast with other papers about voting blocs, Ginsburgh & Noury (2008) did not found evidence to assume that results from the ESC reflected political conflicts / friendships.

In addition to the effect of quality and familiarity variables on voting bias, it appears that performance factors also play a role. A variable that often returns in various studies to comparable competitions is the effect of performance order. This suggests that the order of appearance of a participant influences the number of points that that participant will receive. Ginsburgh & Flores (1996), and also Ginsburgh & van Ours (2003), found evidence that in the Queen Elisabeth Musical Competition the order of appearance influences the final ranking in that competition. Haan et al. (2005) and Bruine de Bruin (2005) tested if this effect also existed in the ESC. It turned out that the order of performance was also important for the final ranking in this competition. It did not matter whether points were given immediately after a performance or that each participant only received points after everyone had sung his/her song. The later a participant has to act, the more advantageous it is for the final ranking.
Clerides and Stengos (2006) also discovered other non-quality related factors such as the language of the song, the gender of the artist, the number of artists on stage and the fact that a participant came from the country that is the host country.
Finally, Doosje & Haslam (2005) discovered that there is a reciprocity factor in the number of points to be awarded to participants. If a country A has given many points to a country B in the last 5 years, then country A receives significantly more points from country B in year *t*. Conversely, country C also gives significantly more points to country D in year *t*, when country C has received many points in the last 5 years from country D.

No previous studies have been done that looked at whether all familiarity and performance factors, mentioned in the above literature, have a significant effect when they are included in one regression analysis. In addition, no research has yet been done into the effect of these factors after the introduction of the new voting system in 2009. In this paper all the above factors will be taken into account and the data will be extended up to and including 2018. The first hypothesis is therefore as follows:

Hypothesis 1:

H_0: There is no effect of familiarity factors on the voting bias
H_1: There is an effect of familiarity factors on the voting bias

The results of testing this hypothesis will make clear which factors affect the voting bias and what is needed to investigate whether those factors can be reduced if the voting system is changed.

Haan et al. (2005) conducted research into the effect of jury type on the voting bias. They have examined the effect of order of appearance on the final ranking and which type of jury was least affected by that non-quality factor. Evidence has been found that a jury of experts recognizes better quality in the sense that they are less sensitive to the order of appearance compared to televoters in national song contests, ceteris paribus. It must be said that the expert jury is still affected by the order of appearance. An explanation given by Haan et al. (2005) is that there is also a performance order effect for jury members, due to the availability heuristic (Tversky & Kahneman, 1973). People can more easily remember candidates who had their performance at the beginning or at the end, which could possibly cause these performances to get higher scores.
Although Haan et al. (2005) already concluded that there was an order effect at the ESC, the data they used did not allow them to distinguish which type of voting system was most influenced by an external factor such as order of appearance. For this reason, this study will also use ESC data from after 1997, to initially examine if the type of voting system influences the size of the voting bias. This will not only look at the jury or experts system and the televoting system, but also at the combination system that has been used since 2009 (see Chapter 2). The second hypothesis is therefore as follows:

Hypothesis 2:

H_0: There is no effect of the type of voting system on the voting bias
H_1: There is an effect of the type of voting system on the voting bias

Secondly, it will be considered which type of voting system recognizes the best quality and is therefore least affected by external factors. The external factors of interest in this study differ from those of Haan et al. (2005). Where they used the order of appearance as an external factor of interest, familiarity factors will be used in this paper to test which voting system recognizes the best quality (more on this in section 4.1).
To test which voting system is most objective, the third hypothesis is the following:

Hypothesis 3:

H_0: There is no effect of familiarity factors on the voting bias, when there is a change of the type of voting system
H_1: There is an effect of familiarity factors on the voting bias, when there is a change of the type of voting system

Testing the hypotheses will show which voting system causes the greatest reduction in the effect of familiarity factors on the voting bias, and therefore also if a change in voting system influences the reduction of familiarity factors at the ESC.

4. Data

To test the hypotheses, the basic data set from the ESC was obtained via the data.world catalog. This dataset consisted of 46,383 observations that provide information about the number of points given by a country to a participant in the period 1975 until 2018. However, in this dataset there were observations for countries that gave themselves 0 points (e.g. Belgium to Belgium). Secondly, the observations for the years 2016-2018 were treated differently, because for the first time, a professional jury and televoting jury of a country independently awarded points to participants.
The observations where countries gave themselves 0 points were removed from the dataset. For the second adjustment of the dataset, the observations, from the jury and televoting points, will be combined to 1 observation per country, instead of 2 observations. This is done by averaging the number of points given by televoters and the number of points given by the professional jury (50% jury points and 50% televoting points). After correcting for these adjustments, 39,432 observations remain.

This dataset is used to calculate if there is a voting bias and what the magnitude is of that voting bias/subjectivity gap. The voting bias is defined based on a formula used by Charron (2013):

(1) $$Bias\ vote_{ijt} = |Vote_{ijt} - Quality_{jt}|$$

This formula explains that the bias equals the absolute difference between the number of points given by country *i* to participant *j* in year *t* and the quality of participant *j* in year *t*. However, because quality is difficult to determine when it comes to music, it has been decided to measure quality based on popularity of the performance (Charron, 2013). It is assumed that the more popular a performance is, the higher the quality will be. The popularity is determined by the total number of points that a participant has received. The quality of the song is also determined by using the formula of Charron (2013):

(2) $$Quality\ vote_{jt} = \frac{1}{p-2} \sum_{p \neq i,j} C_{ijt}$$

This formula means that the quality of participant *j* in year *t* is measured by dividing the total number of points obtained (excluding the points given by country *i* to participant *j*) by the number of countries that give points in a specific competition (i.e. semi-final or a final) minus 2. The formula will give us the most objective number of points a performance should get, according to the majority of the voters. The points from country *i* to participant *j* are not included, because this study looks at the magnitude of the deviation from the number of points given by country *i* compared to quality (i.e. all other countries that give points). So those points should not be included in the sum when quality is measured. The minus 2 is there, because both country *j* itself and country *i* cannot be included in the number of voting countries. This is because country *j* cannot vote for its own participant and the points of country *i* have been taken out. In other words, the voting bias shows the deviation of the number of points given by country *i* compared to the average number of points given by all other countries.

4.1 Familiarity factors

A number of familiarity variables will be used to test whether they have an effect on the voting bias of formula (1). This involves the following variables: Power Distance, Individualism, Masculinity, Uncertainty Avoidance, Long Term Orientation, Indulgence, Shared language, Shared religion, Shared border and Minority score.

The first six variables are the cultural dimensions developed by Hofstede (2011) and used by Ginsburgh & Noury (2008), Spierdijk & Vellekoop (2009) and Budzinski & Pannicke (2017). These dimensions show cultural differences between countries. Countries receive a certain index score for the extent to which they meet the specific dimension. These scores are obtained via Hofstede's country comparison tool (Hofstede Insights, 2019). The country comparison tool contains cultural scores which are determined on the basis of survey data collected by Hofstede from 1966-2010. For the reason that these scores are based on more than 50 years of data, it is assumed in this paper that these scores are constant over time. The range of the index has a minimum value of 0 and a maximum value of 100. The scores are calculated by index specific formulas where results of particular questions from the survey are multiplied by constant factors (Hofstede & Minkov, 2013).
The Power Distance (PDI) measures the extent to which a society accepts that power is unevenly distributed. A score of 100 means that people in society accept their position in the hierarchy, a score of 0 means that people in society strive for an equal distribution of power.
The Individualism score (IDV) shows to what extent there is an expectation that people will take care of themselves (and direct family members). A score of 100 here means an individualistic society and a score of 0 means a collective society.
The Masculinity score (MAS) shows the extent to which society is competitive. A score of 100 shows a preference for achievements and material rewards for those achievements. The opposite, femininity, shows a society that is cooperative, will take care for the weak and wants a high quality of life.
The Uncertainty Avoidance factor (UAI) gives a high score when a society feels uncomfortable with a high degree of uncertainty and ambiguity. For a society with a high score it is all about principles and intolerance of unorthodox ideas or behaviors. A low score shows that a society is mainly concerned with practice, and not sticking so much to principles or traditions.
A score of 100 on Long Term Orientation (LTO) means that companies, for example, focus on good education in order to prepare optimally for the future. A score of 0 shows that companies focus primarily on preserving traditions and looking at changes with suspicion.
Finally, the Indulgence factor (IND), which shows to what extent a society allows people to meet their needs, enjoy life, and have fun. A score of 100 indicates a free society and a score of 0 indicates a society with strict social norms.

The next familiarity factor is a language that is spoken in a country. Only when a language is the official language of a country (unless indicated otherwise), is it considered as the language of a country (Clerides & Stengos, 2006).

The data showed that a country could have a minimum of 1 and a maximum of 4 official languages. If one of the languages of a participant matches the language of the voting country, it is a shared language.
The variable Shared religion is created in the same way. When one of the religions of the participant matches one of the religions of the voting country, then it is a shared religion (Clerides & Stengos, 2006). If more than 5% of people have the same religion in a country, it is considered a relevant religion in a country, where a country can have a maximum of 3 religions. No distinction has been made between movements within Christianity or. In addition, atheism is not considered as religion (missing value).
For the Shared border factor, only “land” borders and not sea borders are taken into account (Spierdijk & Vellekoop, 2009).
Information about language, religion and borders was obtained via CIA World Factbook (2019).
For the variable 'Minority score' it was determined which minority groups of the participant's country live in the voting country. This is a ranking of the 5 largest groups, with the largest minority group receiving the highest score (Charron, 2013). The data on this have been obtained from the paper of Charron (2013) and supplemented with data of CIA World Factbook (2019).
Data on Serbia & Montenegro was no longer available and so it is decided to use Serbia's data on language, religion, minority score for Serbia & Montenegro. For Yugoslavia, the data was also no longer available and therefore data is obtained via the paper of Shoup (1963).

4.2 Voting systems

In addition to familiarity factors, the voting system variable is also relevant to this research. Three types of voting systems have been used at the ESC over the years. From 1975 to 1997, a professional national jury was used to distribute points (Eurovision, Stockholm 1975, 2019). This jury judged participants according to the Borda voting system (Merill, 1984). The jury members rank the participants, whereby the number 1 receives the most points and the last one receives the least number of points. The ESC used a modified version. As mentioned in chapter 2, the number one gets 12 points, number two gets 10 points, number three gets 8 points and numbers four up to ten gets 7 down to 1 point respectively. All participants lower than rank ten receive 0 points.
From 1998 to 2008 and the semifinals of 2009, the televoting system was used (Eurovision, Birmingham 1998, 2019). The televoting jury votes according to the plurality system (Merrill, 1984). Televoters can vote for one participant, who is their favorite. All these votes are then added, which leads to an aggregated ranking.
The last system is the combination of both voting systems and was introduced in the final of the 2009 contest (Royston, 2008). In this system the number of points that is given to a participant consists of 50% of the opinion of the professional jury and 50% of the opinion of the televoters from a country.
Another way to create a variable of the voting system is by looking at the actual voting system used by a country, instead of the Official voting system.

There were a number of countries that have deviated from the officially used system. For example, Norway was using a professional jury in the 2009 final instead of a combination of both. Therefore two voting system methods will be used: the Official voting system method and a country specific voting system method. The Official voting system is already included in the base dataset as a variable, but will indirectly be used, because of collinearity with the Year effects. The Country voting system is created by collecting data manually about the voting system used by a country itself. This was done by viewing the individual events 1975-2018 via the official website of the ESC (Eurovision History, 2019) and via the website that has collected data about those individual events (Eurovision World, 2019). After this data was obtained a dummy variable was created that shows whether there is a deviation of a Country voting system from the Official voting system. So this ensures that the Official voting system method is represented in the deviation variable.

4.3 Control factors

Finally, a number of variables will be included in the dataset that serve as controls. These are the following variables: Order of appearance, Language of the song, Gender of the (lead)singer, Number of artists on stage, Number of voters in (semi-)final, Host country, Reputation, Reciprocity and Lagged vote.

As mentioned in chapter 2, various studies have shown that the order of appearance affects the number of points that a participant receives (Haan et al., 2005; Bruine de Bruin, 2005). It is therefore expected that the order of appearance also affects the voting bias. The data about the Order of appearance was obtained via the catalog of data.world and showed the performance order per participant per year in a (semi-)final.
Next to this, the language of the song, the number of people on stage and the gender of the lead singer is taken into consideration. Following Ginsburgh and Noury (2008), this paper also uses song covers and song titles to create these variables. Based on these song covers it is manually determined whether there was a solo act, a duet or a group/band that gave the performance (Eurovision Covers, 2019). If this was not clear, it was checked via participant lists (Eurovision, History, 2019). It is assumed that a single name is a solo act, two names is a duet and more names or a word etc. is labeled as a group.
Based on the song title it is manually determined if a song was either English, French or German and were they labelled as such in the dataset. These three languages have been chosen, because these three languages are most commonly spoken in addition to their own mother tongue in Europe (European Commission, 2012).
The period from 1977 to 1998 is an exception to the above method. Countries were then obliged to sing in (one of) their own language(s). During this period, the language of the song is automatically labeled as the official language spoken in a country. For example, the participation of Spain in 1980 is labeled as “Spanish”, as that is their official language. If a country has more official languages, then again the label will be based on the song title.
There are some cases that deviate from this rule; Germany and Belgium in 1977 and Poland in 1994.

In 1977, Germany and Belgium had already selected their participants before the rule was introduced and it was too complicated to change, so they received dispensation (Eurovision, London 1977, 2019). Poland had sung an English song against the rules in 1994, but was nevertheless not disqualified by the EBU member states (Eurovision, Dublin 1994, 2019).
The song covers also determine whether the (lead) singer is a man or a woman. If this was unclear, it was observed via video clips of the performance and manually complemented if necessary. A singer was coded as a woman or man when only a man or woman was singing. For example, a duet between man and woman and a group in which both a man and a woman were singing is coded as missing value.
The number of voters is also taken into account for verification. Charron (2013) added the number of participants, because this number has increased significantly since the accession of Central and Eastern European countries to the contest. In this paper a distinction is made between the number of participants and the number of voters in a (semi-) final in year *t*, because this can differ. Only the number of voters is relevant for this paper, as this research is about the voting bias. The voting bias is based on the number of points a participating country received by a *voting* country, so it is assumed that only the number of voting countries is important.
The Host country factor shows the effect of last year's winner (Ginsburgh & Noury, 2008). There is one exception where the factor does not accurately reflect this and that is in 1980. The Netherlands organized the song festival then, but Israel was the winner in 1979. However, they withdrew for the 1980 edition (Eurovision, The Hague 1980, 2019).
The Reputation variable was created by including the number of times a participant had won the contest until year *t*. For example, Ukraine had won the contest in 2004, so from 2005 it had reputation score of 1. In 2016 they won again and so from 2017 the reputation factor gives a score of 2. In 1975 Sweden already had a reputation of 1, because they had won in 1974. The expectation is that countries will rate participants of a country with a high reputation differently than when they have a low reputation score. For instance, out of jealousy, countries can give lower number of points to a participating country with a high reputation score. Or they give the participating country more points, because they think the participant meets the expectations.

The variable Reciprocity means that the number of points that country B has given to participant A in year *t-1* is taken into account in the number of points to be given by country A to participant B in year *t*. Just like in Ginsburgh and Noury (2008), only for data from 1975-2003 this variable is created, since semi-finals were introduced in 2004. The introduction of semi-finals ensured that some countries do not participate in the final every year and therefore no reciprocity can take place. This means that for data from 2004 and onwards there are missing values.
Finally, the Lagged vote factor. Charron (2013) found that in addition to the Reciprocity factor, this Lagged vote also had a significant effect on the number of points to be awarded by country A to participant B. This factor means that the number of points given by country A to participant B in year *t-1* is taken into account in the number of points to be given by country A to participant B in year *t*. For the same reason as for the Reciprocity factor, only for data from 1975-2003 a lagged vote value is created.

5. Methodology

5.1 Hypothesis 1

First of all, Hypothesis 1 is repeated, to subsequently translate this hypothesis into an econometric model, which will be estimated.
The first hypothesis was as follows:

Hypothesis 1:

H_0: There is no effect of familiarity factors on the voting bias
H_1: There is an effect of familiarity factors on the voting bias

To answer this hypothesis the following random effects model is used:

$$(3) \quad Voting\ bias_{ijt} = \beta_0 + \beta_1(familiarity_{ijt}) + \beta_2(controls_{ijt}) + \alpha_i + u_{ijt}$$

The dependent variable is the difference between the number of points participant *j* got from country *i* in year *t* minus the quality of the song of participant *j* in year *t*, as described in formula (1).
The term 'familiarity' covers several familiarity variables as defined in section 4.1. The first six variables (Power Distance, Individualism, Masculinity, Uncertainty Avoidance, Long Term Orientation and Indulgence) are the root squared differences between cultural scores of voting country A and participating country B. It is assumed that the bigger the difference is, the greater the cultural distance is, between those countries for that particular factor.
Next familiarity variable is 'Shared language', where 1 means that voting country and participating country share an official language and 0 if they don't. The same method is used for 'Shared religion' and for 'Shared border'.
The Minority score is, as stated in section 4.1, a representation of the size of the group of immigrants from the participating country living in the voting country. A score of 5 indicates that immigrants originally from the participating country, are the largest minority group in the voting country, a score 4 indicates that immigrants are the second largest minority group in the voting country etc. A score of 0 shows that there is no minority group of the participating country in the voting country or that it is not among the top 5. The top 5 has been compiled by only including countries that were represented at the ESC. If the largest minority group originally is from a country that has never participated in the ESC, and of the second largest minority group is from a country that has participated in the ESC, then the latter group will be considered the largest.

The control variables are factors as described in section 4.3. The Order of appearance is numeric value that indicates when a performance was given; 1 is the first performance, 2 as second, 3 is the third performance etc.

The language of the song is split into three variables, namely an English song, a German song and a French song. A value of 1 indicates that the song was English/German/French, a 0 indicates that the song was sung in another language then one of those three.
The variable 'Male singer' takes a value of 1 if the (lead)singer was a male and value 0 if the (lead)singer was a female.
Number of artists on stage is also split into more variables, namely Duet and Group. A value of 1 for Duet indicates that the song was performed by two people. A value of 1 for Group indicates there was a group/band on stage. If both variables have value 0, then the performance is a solo act.
The variable Number of voters represent the total number of voters in a (semi-)final.
Host country is a dummy variable that has a value of 1 if the country was the hosting country and 0 if the country wasn't.
The Reputation variable is the number of times a participating country had won the contest before year t.
The last two control variables are lagged variables, the Reciprocity variable and the Lagged vote variable. The Reciprocity factor is the number of points the voting country *received* in year *t-1* from the participating country. The Lagged vote is the number of points the voting country *gave* to the participating country in year *t-1*.
Finally, α_i is an unobserved time-invariant country specific error term.

A random effects model is used to estimate the effect of the variables on the voting bias, because the dataset is panel data with some time invariant variables included. In a fixed effects model these time invariant variables would be absorbed by the intercept. Time invariant variables are religion, border and language factors, but also some control variables.
To reduce the effect of heteroskedastic residuals and to control for the effect of intra-country correlation, clustered robust standard errors are used.
Lastly, to check for non-linearity a random effects tobit model is used as a robustness check.

5.2 Hypothesis 2

The second hypothesis to test was as follows:

Hypothesis 2:

H_0: There is no effect of the type of voting system on the voting bias
H_1: There is an effect of the type of voting system on the voting bias

To see what the effect of a voting system is on the voting bias, two voting system variables are used. To start, the effect of an country specific voting system on the voting bias is examined and secondly the deviation from the Official voting system is taken into account.

The model that will be used is to test the effects is as follows:

$$(4)\quad Voting\ bias_{ijt} = \beta_0 + \beta_1(voting\ system_{ijt}) + \beta_2(dev\ off\ voting\ system_{ijt}) + \beta_3(familiarity_{ijt}) + \beta_4(controls_{ijt}) + \alpha_i + u_{ijt}$$

In this model the same dependent variable, familiarity factors and control variables are used as in section 5.1 (formula 3), with the exception of the variables Lagged vote and Reciprocity. These have been dropped out because these variables only contain data from 1975 - 2003. For testing this hypothesis the combination voting system of 2009 has also to be taken into account. So it has been decided to exclude these two variables from the model, in order to increase the number of observations that contains the most recent voting system.

The voting system variable can take a value of 1, 2 or 3. Value 1 means the jury voting system. Value 2 means the televoting system. Value 3 means the combination voting system.
The second way to describe the variable voting system is by defining which voting system is officially had to be used in year *t*. To control for a potential effect of deviation from the Official voting system on the voting bias, this data also will be used. The Official voting system variable can take value 1, 2 or 3 as well. Value 1 means the voting system from 1975-1997 (jury voting), value 2 means the voting system from 1998-2008/2009 and value 3 means the voting system from finals 2009-2018. The deviation variable is created by observing whether there is a difference between the Country voting system and the Official voting system. If there is a difference, the deviation variable takes value 1 and if there is not a difference it will take value 0.

To estimate the effect of the variables on the voting bias also here a random effects model is used. For the same reasons as mentioned in section 5.1 clustered robust standard errors are used. A random effects tobit model is used again as a robustness check for non-linearity.

5.3 Hypothesis 3

The last hypothesis tested in this paper was as follows:

Hypothesis 3:

> *H_0: There is no effect of familiarity factors on the voting bias, when there is a change of the type of voting system*
> *H_1: There is an effect of familiarity factors on the voting bias, when there is a change of the type of voting system*

The model that is used to test this hypothesis is in fact a combination of the models which have been shown section 5.1 (formula 3) and 5.2 (formula 4). only in this model interaction effects are included:

$$(5)\ Voting\ bias_{ijt} = \beta_0 + \beta_1(voting\ system_{ijt}) + \beta_2(dev\ off\ voting\ system_{ijt}) + \beta_3(familiarity_{ijt}) + \beta_4(familiarity_{ijt} * voting\ system_{ijt}) + \beta_5(controls_{ijt}) + \alpha_i + u_{it}$$

All the variables used in the previous two models are used again (formula 3 and 4), with the exception of Reciprocity and Lagged vote. The new variables added to this model are interaction factors, which are interactions between all familiarity variables and the voting system variable. These interaction variables will show which voting system ensures the lowest impact of familiarity variables on the dependent variable (Voting bias). This will therefore show which voting system provides the greatest reduction in the effect of external factors such as sharing a border, religion, ethnicity, culture or language.

Again a random effects model is used to test this model, and clustered robust standard errors are taken into account.

6. Results

In this section first of all the results of the models to test Hypothesis 1 will be described, after this the results of models to test Hypothesis 2 will be shown and lastly the results of the model to test Hypothesis 3 will be discussed. The dataset consists of data from 1975-2018, of 52 different countries. It is an unbalanced dataset, as every country does not participate/vote every year. The number of participants varies between 15 and 28 in a (semi-)final and the number of voters vary between 18 and 43 in a (semi-)final. The dataset consists of 25 variables in total, 12 control variables included. For some variables there are missing values, for instance the cultural factors of Hofstede. Cultural data from Hofstede is available for 38 countries at the moment. For the cultural variable Indulgence only data of 37 countries is available, as there is no information about this factor from Israel. The summary statistics of all variables of interest and controls are shown in Appendix A.

6.1 Familiarity factor model

Table 6.1 shows the familiarity factor models. Model 1 is a GLS random effects model and contains only the familiarity variables, because only these variables are interesting for testing Hypothesis 1. The sample consists of 25,122 observations from 37 countries. Model 2 is also a GLS random effects model where control variables are included and consists of 22.568 observations from 37 countries. Model 3 is the last GLS random effects model in this section and contains again the familiarity factors and controls. The control variables Lagged vote and Reciprocity are only included in this model, because this reduces the number of observations significantly. The sample consists of 6,366 observations from 29 countries from the years 1976-2003. Model 4 is a random effects tobit model to control for non-linearities and the sample consists again of 6,366 observations from 29 countries from the years 1976-2003.

It can be seen that sharing a border gives a significant effect of 0.428 points on the voting bias, ceteris paribus, compared to not sharing a border. This effect is significant at the 1% level in Model 1 and 2. This means that if voting country A and participating country B share a border, country A deviates on average 0.428 points from the quality vote on country B, compared to two countries that are not neighboring countries, ceteris paribus. This result is consistent with Spierdijk and Vellekoop (2009), who also found a positive bias when countries are neighboring countries. The effect of sharing a border is smaller and not significant at the 10% level anymore, when variables Lagged vote and Reciprocity are included, ceteris paribus.

If two countries have a common official language this will have a negative effect of 0.194 points on the voting bias, ceteris paribus. This effect is not significant at the 10% level in Model 1 and 2. However, when we have a look at Model 3, the effect of sharing a language is significant at the 1% level. The effect of -0.354 points means that voting country A deviates on average 0.354 from the quality vote on country B, compared to two countries who do not share an official language, ceteris paribus. This finding is contradicting several studies about voting blocs (Yair, 1995; Clerides and Stengos, 2006; Gatherer, 2006), but is consistent with Ginsburgh and Noury (2008). The reason for this is that Ginsburgh and Noury (2008) also included a quality factor in their analysis, while previous studies did not take quality into account.
Sharing a main religion does have a negative effect, ceteris paribus, compared to not sharing a main religion. The effect of -0.211 is only significant when Lagged vote and Reciprocity is not taken into account at a 5% significance level (Model 2). The effect means that if voting country A and participating country B both have the same main religion, country A is less subjective and their vote is on average 0.211 points closer to the quality vote on country B, ceteris paribus. This result is contradicting the papers of Spierdijk and Vellekoop (2009) & Clerides and Stengos (2006), but as discussed earlier this may have been due to the fact that quality was not included in both studies.

The factor Minority score shows a significant effect on the voting bias in each of the models, at the 1% level. For Model 2 this is a positive effect of 0.271 points, ceteris paribus and for Model 3 this is a positive effect of 0.131 points, ceteris paribus. This factor tells that the larger the minority group of participating country B is in voting country A, the larger the gap between the vote from country A to country B and the quality vote to country B on average, ceteris paribus. This gap can increase to 1.355 points (and 0.655 points for Model 3), as the maximum minority score can be 5. Charron (2013) also found a significant positive bias of Minority score, where the magnitude of the effect was even greater.

As discussed in section 4.1, countries can have an index score between 0 and 100. This means that the difference between the voting country and the participating country can also vary from 0 to 100. An increase of 1 score point means that two countries, culturally speaking, differ little from each other for a particular factor.
For the difference in Power Distance scores, this means that an increase in difference of 1 score point has significant effect of -0.004 points on the voting bias for Model 2, ceteris paribus. This effect is significant at the 1% level. Model 3 does not show a significant effect of PDI difference on the voting bias at the 10% significance level.
The difference in Individualism scores does not show a significant effect in any model, at the 10% significance level.
Masculinity score differences do have a significant effect of -0.003 on the voting bias in Model 2, ceteris paribus. This effect is significant at the 1% level. It shows that the more voting country A and participating country B differ on masculinity index scores, the smaller the gap between the number of points country A assigns to country B and the quality vote of country B on average is, ceteris paribus. For example, a difference of 20 masculinity points, gives a smaller voting bias of 0.06 points, ceteris paribus.

The difference in Uncertainty Avoidance score does not show significant results on the voting bias as well, at the 10% significance level.
Long Term Orientation differences have a significant effect of -0.004 points on the voting bias in Model 2, ceteris paribus. This effect is significant at the 5% level. In Model 3 this effect is even more negative, which is -0.008 points on the voting bias, ceteris paribus, at the 1% significance level. An effect of -0.008 means that the deviation of the vote from voting country A to participating country B compared to the quality vote of country B, decreases with on average 0.008 points when countries are one Long Term Orientation score point more different from each other, ceteris paribus.
Finally the factor Indulgence difference. This factor does not show a significant result on the voting bias, at the 10% significance level.
The effects of the cultural differences on the voting bias are consistent with previous studies. Both Spierdijk and Vellekoop (2009) & Ginsburgh and Noury (2008) show that the effect is negative, but very small per index point. Therefore it can be said that countries that have more similarities, from a cultural point of view, are more inclined to judge more subjectively, since the voting bias then becomes larger, ceteris paribus.

Model 4 is a random-effects tobit model, and is used as a robustness check for Model 3. The results do not significantly deviate from the results of the GLS random effects model. For this reason the results of Model 4 will not be discussed in this section.
For Stata output of all models from this section, see Appendix B.

As stated in section 5.1, in all models of Table 6.1 clustered robust standard errors are used to reduce the potential effect of heteroskedastic residuals and intra-country correlation. Our models also do not suffer from multicollinearity, which appeared from the correlation matrix.

Table 6.1: The impact of familiarity factors on the voting bias.

VARIABLES	Model 1	Model 2	Model 3	Model 4
Shared border	0.432***	0.428***	0.203	0.214*
	(0.138)	(0.144)	(0.126)	(0.120)
Shared language	-0.199	-0.194	-0.354***	-0.346**
	(0.138)	(0.122)	(0.122)	(0.154)
Shared religion	-0.068	-0.211**	-0.061	-0.050
	(0.085)	(0.098)	(0.058)	(0.080)
Minority score	0.252***	0.271***	0.131***	0.133***
	(0.029)	(0.031)	(0.038)	(0.038)
PDI score difference	-0.004***	-0.004***	-0.002	-0.002
	(0.001)	(0.001)	(0.002)	(0.003)
IDV score difference	-0.001	-0.000	-0.003	-0.002
	(0.002)	(0.001)	(0.002)	(0.002)
MAS score difference	-0.003***	-0.003***	-0.005***	-0.005**
	(0.001)	(0.001)	(0.002)	(0.002)
UAI score difference	0.000	-0.001	-0.000	0.000
	(0.001)	(0.001)	(0.003)	(0.003)
LTO score difference	-0.003*	-0.004**	-0.008***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.003)
IND score difference	0.001	0.001	0.002	0.002
	(0.002)	(0.002)	(0.003)	(0.003)
Order of appearance		0.019***	0.024***	0.026***
		(0.002)	(0.004)	(0.004)
Host country		0.328***	0.365***	0.397***
		(0.061)	(0.112)	(0.090)
Reputation		0.008	0.060**	0.056***
		(0.015)	(0.025)	(0.021)
English song		0.149***	0.495***	0.516***
		(0.055)	(0.097)	(0.092)
French song		-0.025	0.355***	0.395***
		(0.067)	(0.083)	(0.083)
German song		-0.075	0.226***	0.167**
		(0.066)	(0.078)	(0.079)
Male singer		-0.007	-0.293***	-0.336***
		(0.028)	(0.044)	(0.048)
Duet		0.173***	0.035	0.076
		(0.053)	(0.101)	(0.082)
Group		-0.092**	-0.013	0.025
		(0.039)	(0.048)	(0.048)
Reciprocity			0.005	0.005
			(0.007)	(0.008)
Lagged vote			0.025***	0.027***
			(0.007)	(0.008)
Constant	2.672***	2.820***	2.123***	2.060***
	(0.163)	(0.177)	(0.157)	(0.162)
Observations	25,122	22,568	6,366	6,366
Number of voting countries	37	37	29	29
Year FE	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Model 1 is the base model, Model 2 control variables are included, with exception of the variables Lagged vote and Reciprocity, Model 3 all control variables are included and Model 4 is the random effects tobit model. Year fixed effects are included in all models, but omitted from presentation in Table 6.1.

6.2 Voting system model

Table 6.2 gives the results of the models where voting system variables are included. Models 5 and 6 are both GLS random effects models and contain 22,568 observations from 37 countries from 1975-2018. The difference between Model 5 and 6 is that Model 6 controls for a potential effect of the deviation variable on the voting bias. Model 7 is a random effects tobit model to control for non-linearities and also consists of 22,568 observations from 37 countries from 1975-2018.

When voting system 2 is used, this has a significant effect of -0.703 (-0.516) points on the voting bias in Model 5 (Model 6), compared to using voting system 1, ceteris paribus. This effect is significant at the 1% level. This means that the deviation of the number of points that voting country A gives to participating country B compared to the quality vote decreases on average when a televoting jury is used, ceteris paribus.
The effect of using voting system 3 has also a significant negative effect of 0.364 (0.245) points on the voting bias in Model 5 (Model 6), compared to voting system 1, ceteris paribus. Only this effect is significant at the 5% significance level. The gap between the number of points given by voting country A to participating country B and the quality vote to country B will decrease on average as well as the combination voting system is used, ceteris paribus.
These results tell us that voting system 1 has the greatest bias compared to other voting systems, ceteris paribus. This contradicts the results of Haan et al. (2005), as they found that a jury of experts is less prone to external factors compared to a televoting system. However, they did research in national song contests, which is different in terms of external influences. External factors like for instance common languages and common borders do not play a role in national song contests. In addition, Haan et al. (2005) only looked at the effect of order or appearance on the final ranking and whether this differs with different types of voting systems. In this study, the Order of appearance has been included to control and it can be seen that a voting bias remains, which is larger in a jury voting system compared to the other two systems.

In Model 6 the deviation variable shows a significant result of 0.271 points on the voting bias, ceteris paribus. This effect is significant at the 1% level. It shows that the voting bias increases on average when a voting country uses a different voting system than the one that should have been used officially, ceteris paribus. This effect can partly be explained by the fact that 1286 out of 1735 observations, that deviate from the Official voting system, use the jury voting system instead of the televoting system or combination system. Using a jury voting system is more biased than the other two systems, ceteris paribus, so for that reason it partly explains the increase in the voting bias.

The familiarity factors do not show significant differences compared to the models of section 6.1 and therefore these factors will not be discussed in this section.
Also Model 7, which is the random effects tobit model, does not show significant differences with Model 6. So these results will also not be discussed in this section.
For Stata output of all models from this section, see Appendix C.

There is also checked for the variables of Model 5-7 if they suffer from multicollinearity. The correlation matrix shows that there is no multicollinearity. Finally, as mentioned in section 5.2, cluster robust standard errors are used to reduce the potential effect of heteroskedasticity.

Table 6.2: The impact of voting system variables and familiarity factors on the voting bias.

VARIABLES	Model 5	Model 6	Model 7
Country voting system 2	-0.703***	-0.516***	-0.543***
	(0.140)	(0.116)	(0.198)
Country voting system 3	-0.364**	-0.245**	-0.215
	(0.145)	(0.118)	(0.161)
Deviation country voting system		0.271***	0.250
		(0.0943)	(0.162)
Shared border	0.430***	0.430***	0.461***
	(0.145)	(0.145)	(0.156)
Shared language	-0.179	-0.180	-0.205
	(0.123)	(0.123)	(0.133)
Shared religion	-0.173	-0.173	-0.227*
	(0.111)	(0.111)	(0.129)
Minority score	0.272***	0.272***	0.271***
	(0.031)	(0.031)	(0.040)
PDI score difference	-0.004***	-0.004***	-0.004**
	(0.001)	(0.001)	(0.002)
IDV score difference	-0.000	-0.000	0.000
	(0.002)	(0.002)	(0.002)
MAS score difference	-0.003***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)
UAI score difference	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.002)
LTO score difference	-0.004**	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)
IND score difference	0.001	0.001	0.000
	(0.002)	(0.002)	(0.002)
Constant	2.812***	2.806***	2.864***
	(0.175)	(0.175)	(0.246)
Observations	22,568	22,568	22,568
Number of voting countries	37	37	37
Control variables included	YES	YES	YES
Year FE	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: In Model 5 is the base model, in Model 6 the deviation variable is included and Model 7 is the random effects tobit model. Year fixed effects and performance factors are taken into account in all models, but not shown in this table. Two factors are excluded from these models: Lagged vote and Reciprocity. The reason being that these variables contained data up to 2003, which would mean that voting system 3 could not be included in the regression.

6.3 Interaction effects model

In Table 6.3 the results of the most extensive model are shown. This model contains all the variables used in models before, and interaction variables of the familiarity factors and voting systems. Model 8 is also a GLS random effects model and consists of 22,568 observations from 37 countries from 1975-2018.

Table 6.3: The impact of voting system variables, familiarity factors and interaction effects on the voting bias.

VARIABLES	Model 8
Country voting system 2	-0.384
	(0.316)
Country voting system 3	-0.411*
	(0.239)
Deviation official voting system	0.317***
	(0.0964)
Shared border	0.120
	(0.120)
Shared border Country voting system 2	0.432
	(0.340)
Shared border Country voting system 3	0.397*
	(0.240)
Shared language	-0.134
	(0.114)
Shared language Country voting system 2	0.028
	(0.255)
Shared language Country voting system 3	0.101
	(0.221)
Shared religion	0.182
	(0.111)
Shared religion Country voting system 2	-0.687**
	(0.304)
Shared religion Country voting system 3	-0.488*
	(0.271)
Minority score	0.050*
	(0.028)
Minority score Country voting system 2	0.477***
	(0.085)
Minority score Country voting system 3	0.198***
	(0.062)
PDI score difference	-0.009***
	(0.002)
PDI score difference Country voting system 2	0.005
	(0.004)
PDI score difference Country voting system 3	0.004
	(0.003)

(Continued)

VARIABLES	Model 8
IDV score difference	-0.006*
	(0.003)
IDV score difference Country voting system 2	0.009
	(0.005)
IDV score difference Country voting system 3	0.007**
	(0.004)
MAS score difference	-0.005***
	(0.002)
MAS score difference Country voting system 2	0.001
	(0.004)
MAS score difference Country voting system 3	0.004
	(0.003)
UAI score difference	0.004
	(0.003)
UAI score difference Country voting system 2	-0.006
	(0.004)
UAI score difference Country voting system 3	-0.005
	(0.003)
LTO score difference	-0.008***
	(0.003)
LTO score difference Country voting system 2	0.003
	(0.004)
LTO score difference Country voting system 3	0.009***
	(0.003)
IND score difference	0.001
	(0.003)
IND score difference Country voting system 2	0.000
	(0.004)
IND score difference Country voting system 3	0.002
	(0.003)
Constant	2.863***
	(0.257)
Observations	22,568
Number of voting countries	37
Control variables included	YES
Year FE	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Year fixed effects and performance factors are taken into account in the models, but not shown in this table. Two factors are excluded again from these models: Lagged vote and Reciprocity. The reason being that these variables contained data up to 2003, which would mean that voting system 3 could not be included in the regression.

In this section only the results that are relevant for the interpretation of the conditional Hypothesis 3 will be discussed. In other words, only the interaction effects will be discussed.

It can be seen that in Table 6.3 a few interaction effects have a significant effect on the voting bias. When voting country A and participating country B share a border, and the combination voting system is used, it will have a significant positive effect on the gap between the number of points country A gives to B and the quality vote of country B. This effect is significant at the 10% significance level.
If two countries have a common main religion, and the televoting system is used, then the effect is negative on the voting bias, with a significance at the 5% level. This is also the case when the combination voting system is used. The effect of having the same main religion, when voting system 3 is used, is negative at the 10% significance level.
The minority score interacting with voting system 2 gives a significant positive effect, at the 1% significance level. This means that, if a televoting system is used and the minority score increases by 1, this will lead to a greater subjectivity of voting country A towards participating country B, when they have to assign points to country B. This also applies when voting system 3 is used. The minority score interacting with the combination voting system will have a significant positive effect, at the 1% significance level.
The interaction between the difference in Individualism score of voting country A & participating country B, and voting system 3, has an significantly positive effect on the voting bias, at the 5% level. If the difference in Individualism scores of those countries increases by 1 index point, using the combination voting system, then this has a positive effect on the gap between number of points country A assigns to country B and the number of points to country B that one would expect if points were assigned on quality basis alone.
Finally, the interaction effect between differences Long Term Orientation scores and voting system 3 has a significant positive effect on the voting bias, at the 1% level. This means that if voting country A uses a combination voting system and the difference in LTO scores increases by 1 index point, this will have a positive effect on the gap between the number of points of country A to country B and the quality vote of country B.
As mentioned in section 5.2, cluster robust standard errors are used to reduce the potential effect of heteroskedasticity.
The Stata output about these results and results of variables which have not been discussed in this section, can be seen in Appendix D.

For the reason that the magnitude of the interaction effects cannot be estimated from Model 8, the average marginal effects of the familiarity variables for each of the different voting systems were examined. In addition, it was examined whether a change in the voting system will result in a reduction/an increase in the effect of the relevant familiarity factor on the voting bias. The Stata output of the average marginal effects can again be found in Appendix D.

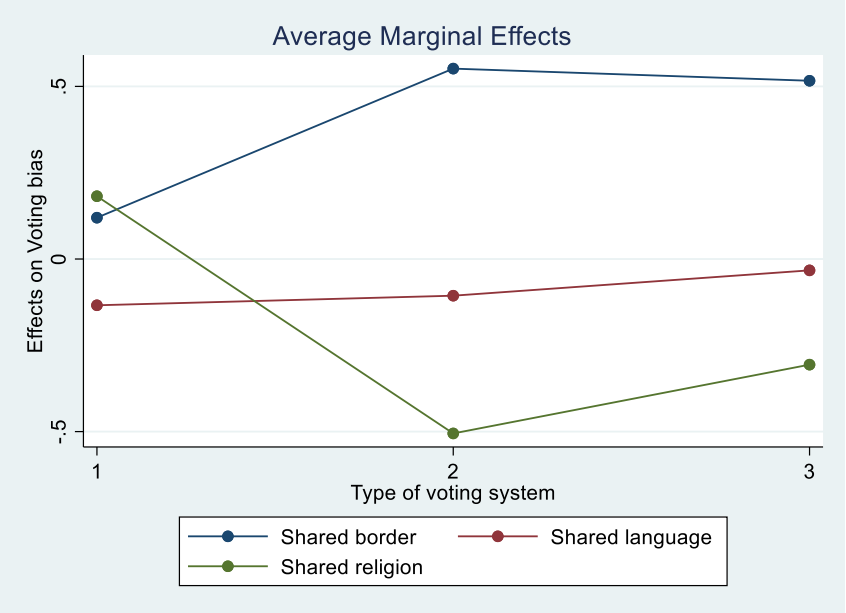

Figure 6.1. Average marginal effects of Shared border, Shared language and Shared religion on Voting bias.

In Figure 6.1 the average marginal effects of Shared border, Shared language and Shared religion are shown. The average marginal effect of sharing a border, when a jury voting system is used, is 0.120 points on the voting bias, ceteris paribus. However, this positive effect is not significant at the 10% level. The average marginal effect of sharing a border when a televoting system is used, is significantly positive, at the 10% level. This effect of 0.552 points shows that if two countries share a border and voting system 2 is used, this will increase the voting bias on average by 0.552 points, ceteris paribus. Finally, if voting system 3 is used, then the effect of sharing a border is on average 0.516 points on the voting bias, ceteris paribus. This effect is significant at the 1% level.
Post-estimations have been done to determine whether a change in voting system causes a reduction / increase in the 'Shared border' effect on the voting bias. These post-estimations test whether the Shared border coefficients per voting system differ significantly from each other and how large this difference is. The post-estimations show that only the coefficients of voting system 1 and 3 for the variable Shared border, significantly differ from each other. The difference is on average 0.397 points on the voting bias and is significant at the 10% level, ceteris paribus. This means that switching from voting system 1 to 3 increases the average marginal effect that sharing a border has on the voting bias significantly, ceteris paribus.

The average marginal effects per voting system for the variable 'Shared language' are all not significantly different from zero. So for this reason, these will not be interpreted any further.

The last variable shown in Figure 6.1 is ‘Shared religion’. If the voting country and the participating country have a common main religion, and a jury voting system is used, this has an effect of on average 0.182 points on the voting bias, ceteris paribus. However, this average marginal effect has a p-value of 0.103 and is therefore not significant at the 10% level. When a televoting system is used then the average marginal effect of sharing a main religion does have a significant effect on the voting bias, ceteris paribus. The average marginal effect is -0.505 points on the voting bias, ceteris paribus, and is significant at the 5% level. Finally, using the combination system also shows that sharing a religion has a negative effect on the voting bias. This average marginal effect has p-value of 0.103 and is, like the average marginal effect of ‘Shared religion’ when voting system 1 is used, not significant at the 10% level.
When differences between the Shared religion coefficients are tested for being equal, the results show that the coefficients of voting system 1 and 2 significantly differ from each other. This difference is 0.687 points and significant at the 5% level. Next to this, the difference between coefficients of voting system 1 and 3 also significantly differ from each other. This difference is 0.488 points and significant at the 10% level.
From these results it can therefore be concluded that changing from a jury voting system to a televoting system or combination system results in a significant reduction of the average marginal effect of 'Shared religion' on the subjectivity of voting country A in assigning points to participating country B, ceteris paribus.

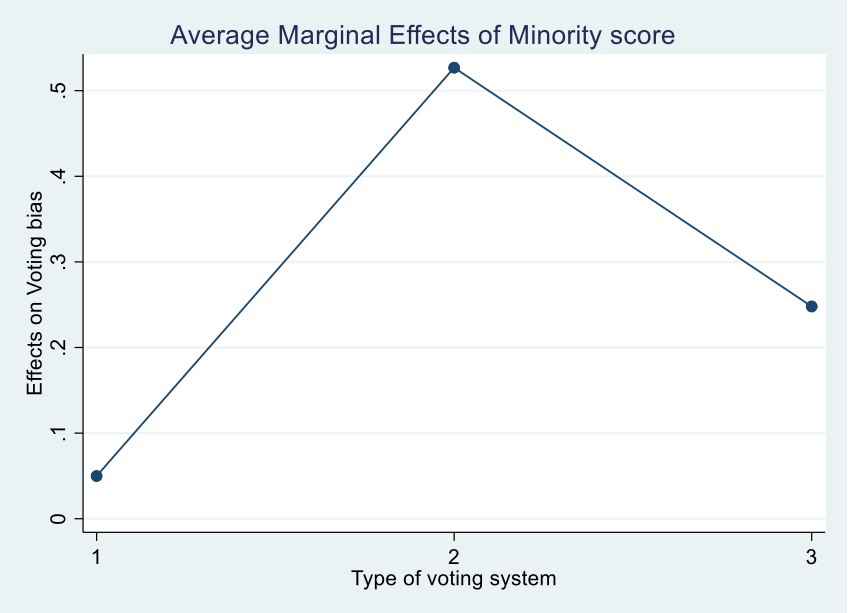

Figure 6.2. Average marginal effects of Minority score on Voting bias.

In Figure 6.2 the average marginal effects of Minority score are shown. When a jury voting system is used, an increase of one minority score point will have a significant positive average marginal effect on the voting bias, ceteris paribus. This effect is on average 0.050 points per minority score point and significant at the 10% level. It shows that the bigger the minority group of participating country B living in voting country is A, the greater the gap between the number of points given by country A to country B and the quality vote is, when voting system 1 is used, ceteris paribus.

If a televoting jury is used, the average marginal effect of Minority score is even greater on the voting bias, ceteris paribus. This is a significant average marginal effect of 0.527 points per minority score point on the voting bias, ceteris paribus, and is significant at the 1% level. Thirdly, when a combination voting system is used, the average marginal effect of Minority score is 0.248 per score point on the voting bias. The average marginal effect is significant at the 1% level.
Tests that show whether the coefficients of these average marginal effects differ significantly from each other, indicate that all coefficients of the Minority score per voting system differ significantly, at the 1% significance level. This means that it can be concluded that a jury voting system on average can best be used to reduce the influence of the Minority score on the subjectivity of voting country A, compared to the other two voting systems, ceteris paribus.

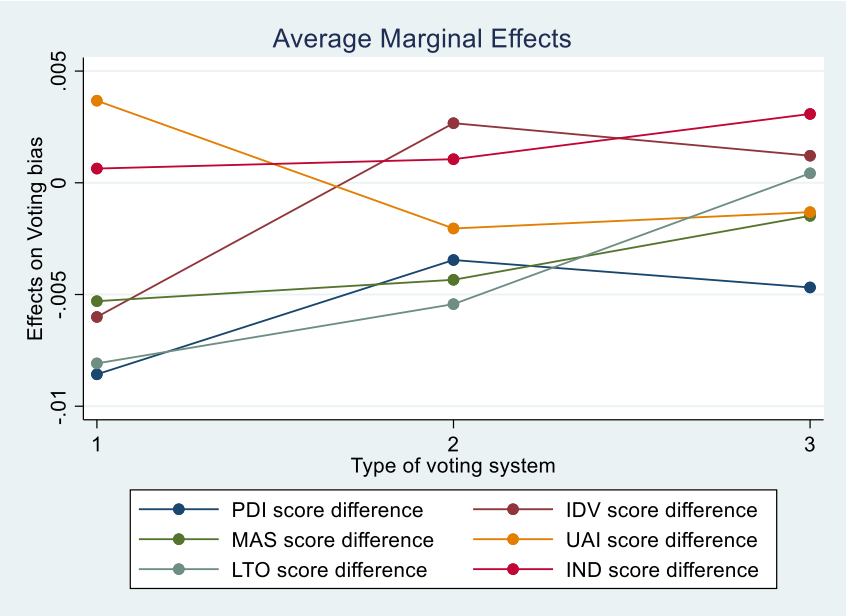

Figure 6.3. Average marginal effects of cultural variables on Voting bias.

Figure 6.3 shows the average marginal effects of all the cultural variables on the voting bias. First variable to discuss is the difference in Power Distance index scores. The average marginal effect of the difference in Power Distance index scores between voting country A and participating country B on the voting bias is significantly negative when voting system 1 is used, ceteris paribus. This effect of -0.009 per index point is significant at the 1% level. The interpretation of this result is as follows: when the cultural difference of Power Distance of voting country A and participating country B increases with one index point, it will lead to an average marginal decrease of 0.008 points of the voting bias, if a jury voting system is used, ceteris paribus. If a televoting jury is used this will lead to a negative average marginal effect of -0.003 points on the voting bias, but is not significant at the 10% level. When the combination voting system is used, this will lead to a significant decrease of the average marginal effect of PDI score difference on the voting bias, ceteris paribus. This effect of -0.005 per index point is significant at the 1% level.

The post estimations did not show a significant difference in coefficients of the average marginal effects of PDI score difference. So it is not possible to determine which voting system can best be chosen to reduce the average marginal effect that the PDI score difference has on the voting bias.

The second variable to discuss is the difference in Individualism scores. Only when voting system 1 is used, the average marginal effect of -0.006 on the voting bias is significant at the 10% level, ceteris paribus. The average marginal effect of the difference in IDV scores on the voting bias when other voting systems are used, did not show a significant effect, at the 10% significance level. This means that if the difference between voting country A and participating country on IDV score increases by one index point, it will affect the voting bias in a negative way with an average marginal effect of -0.006 per index point, when voting system 1 is used, ceteris paribus.
The result of the test of whether the coefficient of the average marginal effect of IDV on voting bias, when voting system 1 is used, significantly differs from the coefficient when voting system 3 is used, shows that there is a difference between them. This difference of 0.007 points is significant at the 5% level. In other words it can therefore be said that it is better to use a jury voting system instead of a combination system, in order to reduce the average marginal effect of IDV score differences on the voting bias, ceteris paribus.

The third cultural variable is the difference in Masculinity scores. If voting system 1 is used, the average marginal effect of the difference in Masculinity scores is -0.005 on the voting bias, ceteris paribus. This effect is significant at the 1% level. It means that if the cultural difference between voting country A and participating country B increases by one index point, this will lead to a decrease in the average marginal effect of MAS score difference on the voting bias, when country A uses voting system 1, ceteris paribus. In other words, the farther apart these countries are in this cultural field, the more the subjectivity of country A on average is reduced, which it has when assigning a number of points to country B and using a jury of experts. When the voting country uses a televoting system the average marginal effect of MAS score difference is -0.004 points on the voting bias, ceteris paribus. This effect is significant at the 10% level. Using the third voting system does not lead to a significant average marginal effect of MAS score difference on the voting bias, at the 10% significance level.
The differences in coefficients of this MAS score difference factor do not show significant results and therefore will not be discussed. So for the factor MAS score difference it cannot be determined which voting system is best to use to decrease the average marginal effect it has on the subjectivity of voting country A towards participating country B.

This also applies to the factors Uncertainty Avoidance score difference and Indulgence score difference. These factors do not show significant average marginal effects on the voting bias, at the 10% significance level. Also, the coefficients of those average marginal effects do not differ significantly from each other. So nothing significant can be said about which voting system would be better or worse to use to reduce the effect of UAI and IND on the voting bias.

The last factor to be discussed in this section is the difference in Long Term Orientation scores. When a jury voting system is used the LTO score difference will have a significant average marginal effect of -0.008 per index point on the voting bias, ceteris paribus. This effect is significant at the 1% level. This result shows that if two countries differ one index point more in Long Term Orientation field, it reduces the gap between number of points of voting country A assigned to participating country B and the quality vote of quality B significantly, when voting system 1 is used, ceteris paribus. When voting system of televoting is used, the average marginal effect of LTO score difference is -0.005 per index point on the voting bias, ceteris paribus. This effect is significant at the 5% significance level. There is no significant average marginal effect of LTO score difference on the voting bias, when the combination voting system is used, because the p-value is 0.814.

The differences in coefficients of the factor LTO score difference do significantly differ from each other. The difference of average marginal effects between using voting system 1 and using voting system 3 is -0.009, at the 1% significance level. And the difference of average marginal effects between using voting system 2 and voting system 3 is -0.006, at the 5% level. This means that changing from a combination system to a jury or televoting system reduces the average marginal effect that LTO score difference has on the voting bias significantly, ceteris paribus.

6.4 Evaluation hypotheses

The hypotheses can be evaluated based on the insights from the above models.

The first hypothesis was as follows:

Hypothesis 1:

H_0: There is no effect of familiarity factors on the voting bias
H_1: There is an effect of familiarity factors on the voting bias

Based on the results of Model 2 and 3 (Table 6.1), the null hypothesis can be rejected at the 1% significance level. The models show that there are a number of factors that affect the voting bias.
A significant effect on the voting bias is observed when: two countries share a border; share a language; share a religion; a big minority group of the participating country lives in the voting country; there is a difference in degree of power distance between two countries; there is difference in degree of masculinity between two countries; there is a difference in degree of long term orientation, ceteris paribus.

The second hypothesis to be evaluated was as follows:

Hypothesis 2:

H_0: There is no effect of the type of voting system on the voting bias
H_1: There is an effect of the type of voting system on the voting bias

When we take a look at the results of Model 5 and 6 (Table 6.2), the null hypothesis can be rejected at the 1% significance level. Shown is that voting system 2 and 3 have a significant negative effect on the voting bias, compared to voting system 1, ceteris paribus. After controlling for the deviation variable, the negative effects were still significant although they were less negative compared to Model 5, ceteris paribus.

The last hypothesis to be evaluated was as follows:

Hypothesis 3:

H_0: There is no effect of familiarity factors on the voting bias, when there is a change of the type of voting system

H_1: There is an effect of familiarity factors on the voting bias, when there is a change of the type of voting system

In Model 8 (Table 6.3) it can already be observed that for some familiarity factors changing the voting system has a significant effect on the voting bias. So it can be concluded that the null hypothesis can be rejected at the 1% significance level.

Based on the average marginal effects results, it can also be concluded that in most cases a jury of experts causes a significant reduction, or at least the smallest significant increase, in the effect that a particular familiarity factor has on the voting bias, ceteris paribus. It is only in order to reduce the effect of having a common main religion on the voting bias, is it better to use a system different than voting system 1, ceteris paribus.

This finding is consistent with Haan et al. (2005), because they have also found that the external factor (order of appearance) was of least influence when a jury of experts is used instead of jury of laypeople (televoting system). They argued that for this reason a jury of experts is better at recognizing quality than the general public. In this study it can also be argued that the effects of familiarity factors have the least impact on the jury of experts, because the voting bias is the smallest. So this means that the experts can judge most objectively.

7. Conclusion

The first contribution of this study has been to examine which familiarity factors affect the voting bias. Previous researches have shown that several geographical and cultural factors have a significant effect on the voting behavior of people, and in particular voting behavior of people at the Eurovision Song Contest. This paper has looked at several factors that have been previously investigated, taken together in one analysis and on top of that extended the number of factors that should be taken into account in the analysis. It is proved that these factors have a significant effect on the voting bias that occurs at the ESC.
In addition, the paper of Haan et al. (2005) has been followed in order to examine if the type of voting system itself has a significant effect on the voting bias. It has been shown that the type of voting system does have a significant effect on the voting bias. If a jury of experts is been used by voting country A this will lead to bigger gap between the number of points this jury of expert of country A assign to participating country B and the quality vote of country B, compared to the other two voting systems. This means that there are still some factors, which have not been controlled for, that affect the jury of experts more than they does affect the jury of laypeople (televoting system).
The final contribution of this paper is, also following Haan et al. (2005), testing to what extent changing the type voting system would be successful in reducing the effect of the familiarity factors on the voting bias. It appears that changing the voting system does indeed have a significant effect on reducing familiarity factors. In contrast with the previous result that a jury of experts is prone to be more biased in recognizing quality, this type of voting system is best to use to reduce the effect of most familiarity factors on the voting bias.
So this means that a jury of experts can best be used when the organization of the ESC wants to reduce the effect of external familiarity factors on the final outcome as much as possible, in order to have in the end a more objective ranking based on quality. In a broader perspective, it can therefore be suggested that experts are less affected by external factors than laypeople, and are therefore better at recognizing quality.

On the other hand, there are also some limitations in this research, which should be taken into consideration in future research. The most important ones are about the definitions of the voting bias factor and the quality factor. In the voting bias formula, the positive and negative deviations cannot now be distinguished from each other. This means that a jury from voting country A can like a German song as much as it dislikes a French song, while the quality vote might be in between. This ensures that the difference between the number of points given to the German song and the quality vote is the same as the difference between the number of points given to the French song and the quality vote. For this reason, the voting bias can therefore give a biased view itself. The suggestion for future research would be to distinguish negative and positive differences and do a separate analysis on both. This would help to see if the effect of external factors on the positive voting bias significantly differ from the effects of external factors on the negative voting bias.

The definition of quality is now based on the average vote of all countries to participant i, excluding the vote of country j. As discussed it is hard to define quality of an artistic creation. For that reason, a popularity formula is now used to indicate quality, because popularity represents a kind of quality. The more people think that a song should receive the highest number of points, the higher the quality will be. In future research it still will be hard to measure quality of an artistic creation, but popularity could be defined in another way. For example, based on the average bookmakers' odds. Bookmakers place bets for every edition of the ESC who should win the song festival. This independent party assesses which song/artist is the most popular on the basis of various indicators and so these quotes could serve as a good indication of popularity. The lower the average odds are, the higher the final ranking, and so the number of points a participant would get.

Next to the definitions of the dependent variable, the definitions of some familiarity factors also could be defined in another way. The variable Shared border is now defined on basis of mainland, so two countries only share a border if they are neighbors via mainland. This could be improved by also take into consideration "sea" borders. For the variable 'Shared religion' the religions could be more specifically labelled. They are now labelled as Christian, Islamic or Jewish, but could be more specified in Catholic, Orthodox, Protestant, Shi'ites, Sunnis etc. If two countries share a language this is now based on having a common official language spoken in a country. In future research this could be based on languages groups, because some languages are officially not the same, but are closely related to each other, see also Dyen, Kruskal, & Black (1992). Further research also needs to been done on Hofstede's cultural factors, as data was not available for all countries which were used in this paper. Adding to this, in this paper it is assumed that these cultural scores were constant over time, but there is a possibility they will change in time. For variables Lagged vote and Reciprocity data also could be extended. In section 6.1 it is shown that the variable Lagged vote had a significant effect on the voting bias, but could not be included in the other models as there was not enough data available. If data of the variables Lagged vote and also Reciprocity can be extended, then they also can be controlled for the potential effect of those factors in the other models like the one's estimated in this paper.

Thirdly I would suggest thinking about more variables which could be influencing the voting bias. In this paper it is assumed that the familiarity factors which are used capture a 'shared taste' and that, based on this taste, countries would vote more subjectively for participating countries who have that taste in common. However, there are plenty of factors that could be capturing a shared taste as well, for instance a factor like a common main sport or the export level between country A and B. It shows that is hard to include all factors that capture a shared taste and that future research still can contribute to express this shared taste even better.

The last suggestion for other researchers is to have a look at the effect of familiarity variables by changing the voting system, on individual country level instead of aggregated level. This might change the effects of some variables, because the magnitude of the individual variables differs per country.

References

Bruine de Bruin, W. (2005). Save the last dance for me: unwanted serial position effects in jury evaluations. *Acta Psychologica*, 245-260.

Budzinski, O., & Pannicke, J. (2017). Culturally biased voting in the Eurovision Song Contest: Do national contests differ? *Journal of Cultural Economics*, 343-378.

Charron, N. (2013). Impartiality, friendship-networks and voting behavior: Evidence from voting patterns in the Eurovision Song Contest. *Social Networks*, 484-497.

CIA. (2019, July 3). *World Factbook*. Retrieved from cia.gov: https://www.cia.gov/library/publications/the-world-factbook/

Clerides, S., & Stengos, T. (2006). Love thy Neighbor, Love thy Kin: Voting Biases in the Eurovision Song Contest.

Dekker, A. (2007, February 19). The Eurovision Song Contest as a 'Friendship' Network. *Connections*, pp. 53-58.

Doosje, B., & Haslam, S. A. (2005). What Have They Done for Us Lately? The Dynamics of Reciprocity in Intergroup Contexts. *Journal of Applied Social Psychology*, 508-535.

Dyen, I., Kruskal, J. B., & Black, P. (1992). An Indoeuropean Classification: A Lexicostatistical Experiment. *American Philosophical Society*, 1-132.

EBU. (2019, July 3). *The largest community of public service media organizations in the world*. Retrieved from ebu.ch: https://www.ebu.ch/about/members

Economist, T. (2013, May 20). *English atop the Eurovision pile, yet again*. Retrieved from Economist.com: https://www.economist.com/johnson/2013/05/20/english-atop-the-eurovision-pile-yet-again

European Commission. (2012). *Special Eurobarometer 386: Europeans and their languages.* Retrieved from http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_386_en.pdf

Eurovision. (2016, October 31). *43 countries to participate in Eurovision 2017*. Retrieved from Eurovision.tv: https://eurovision.tv/story/43-countries-to-participate-in-eurovision-2017

Eurovision. (2019, July 3). *Birmingham 1998*. Retrieved from eurovision.tv: https://eurovision.tv/event/birmingham-1998

Eurovision. (2019, July 3). *Dublin 1994*. Retrieved from eurovision.tv: https://eurovision.tv/event/dublin-1994

Eurovision. (2019, July 3). *History*. Retrieved from eurovision.tv: https://eurovision.tv/events

Eurovision. (2019, July 3). *In a Nutshell*. Retrieved from eurovision.tv: https://eurovision.tv/history/in-a-nutshell

Eurovision. (2019, July 3). *London 1977*. Retrieved from eurovision.tv: https://eurovision.tv/event/london-1977

Eurovision. (2019, July 3). *Moscow 2009*. Retrieved from eurovision.tv: https://eurovision.tv/event/moscow-2009

Eurovision. (2019, July 3). *Rome 1991*. Retrieved from eurovision.tv: https://eurovision.tv/event/rome-1991

Eurovision. (2019, July 3). *Rules*. Retrieved from eurovision.tv: https://eurovision.tv/about/rules

Eurovision. (2019, July 3). *Stockholm 1975*. Retrieved from eurovision.tv: https://eurovision.tv/event/stockholm-1975

Eurovision. (2019, July 3). *The Hague 1980*. Retrieved from eurovision.tv: https://eurovision.tv/event/the-hague-1980

Eurovision. (2019, July 3). *Voting*. Retrieved from eurovision.tv: https://eurovision.tv/about/voting/

Eurovision Covers. (2019, July 3). *Eurovision Song Contest Record Covers*. Retrieved from eurovisioncovers.co.uk: http://www.eurovisioncovers.co.uk/

Eurovision World. (2019, July 3). *Eurovision Voting & Points*. Retrieved from eurovisionworld.com: https://eurovisionworld.com/eurovision

Fenn, D., Suleman, O., Efstathiou, J., & Johnson, N. F. (2006). How does Europe make its mind up? Connections, cliques, and compatibility between countries in the Eurovision Song Contest. *Physica A*, 579-598.

Gatherer, D. (2006). Comparison of Eurovision Song Contest simulation with actual results reveals shifting patterns of collusive voting alliances. *Journal of Artificial Societies and Social Simulation*.

Ginsburgh, V., & Flores, R. G. (1996). The Queen Elisabeth Musical Competition: How Fair is the Final Ranking. *Journal of the Royal Statistical Society*, 97-104.

Ginsburgh, V., & Noury, A. G. (2008). The Eurovision Song Contest. Is voting political or cultural? *European Journal of Political Economy*, 41-52.

Ginsburgh, V., & van Ours, J. C. (2003). Expert Opinion and Compensation: Evidence from a Musical Competition. *THE AMERICAN ECONOMIC REVIEW*, 289-296.

Haan, M. A., Dijkstra, S. G., & Dijkstra, P. T. (2005). Expert Judgment Versus Public Opinion – Evidence from the Eurovision Song Contest. *Journal of Cultural Economics*, 59-78.

Hofstede Insights. (2019, July 3). *Country Comparison*. Retrieved from hofstede-insights.com: https://www.hofstede-insights.com/country-comparison/

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*. doi:https://doi.org/10.9707/2307-0919.1014

Hofstede, G., & Minkov, M. (2013, May). *VSM 2013.* Retrieved from geerthofstede.com: https://geerthofstede.com/research-and-vsm/vsm-2013/

Merill, S. (1984). A Comparison of Efficiency of Multicandidate Electoral Systems. *American Journal of Political Science*, 23-48.

O'Connor, J. K. (2010). *The Eurovision Song Contest: The Official History*. London: Carlton Books Ltd.

Royston, B. (2008, December 8). *EBU confirms 50/50 vote for Eurovision Song Contest*. Retrieved from esctoday.com: http://esctoday.com/12655/ebu_confirms_5050_vote_for_eurovision_song_contest

Shoup, P. (1963). Yugoslavia's national minorities under communism. *Slavic Review*, 64-81.

Spierdijk, L., & Vellekoop, M. (2009). The structure of bias in peer voting systems: lessons from the Eurovision Song Contest. *Empirical Economics*, 403-425.

Spiteri, S. (2019, May 28). *Here are the viewing figures for Eurovision Song Contest 2019*. Retrieved from eurovisionworld.com: https://eurovisionworld.com/esc/here-are-the-viewing-figures-for-eurovision-song-contest-2019

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 207-232.

Yair, G. (1995). 'Unite Unite Europe' The political and cultural structures of Europe as reflected in the Eurovision Song Contest. *Social Networks*, 147-161.

Appendix A

Summary statistics:

Variable		Mean	Std. Dev.	Min	Max	Observations
Year	overall	2004.223	11.21066	1975	2018	N = 39432
	between		7.320567	1980	2016.362	n = 52
	within		9.970263	1973.295	2022.42	T-bar = 758.308
Country ID	overall	27.07509	14.69593	1	52	N = 39432
	between		15.15476	1	52	n = 52
	within		0	27.07509	27.07509	T-bar = 758.308
Participant ID	overall	27.00948	14.73098	1	52	N = 39432
	between		1.038211	24.80782	30	n = 52
	within		14.70889	-.266207	53.20167	T-bar = 758.308
Number of participants	overall	22.77161	3.276072	15	28	N = 39432
	between		.8261433	19	23.73091	n = 52
	within		3.236746	14.36602	28.40765	T-bar = 758.308
Number of voters	overall	30.42491	9.555803	18	43	N = 39432
	between		3.736997	19	37.25899	n = 52
	within		9.167163	13.47218	46.58935	T-bar = 758.308
Voting bias	overall	2.252349	2.016882	0	12	N = 39432
	between		.2298708	1.800584	3.343137	n = 52
	within		2.011625	-.7378471	12.08032	T-bar = 758.308
Shared language	overall	.0876446	.2827809	0	1	N = 39432
	between		.0551584	0	.2294654	n = 52
	within		.2771013	-.1418209	1.081453	T-bar = 758.308
Shared religion	overall	.0462568	.2100435	0	1	N = 39432
	between		.0615222	0	.2697548	n = 52
	within		.2025848	-.2234979	1.044918	T-bar = 758.308
Shared border	overall	.9066241	.2909622	0	1	N = 39432
	between		.1737035	.2869565	1	n = 52
	within		.2433067	-.035822	1.619668	T-bar = 758.308
Minority score	overall	.367468	1.104944	0	5	N = 39432
	between		.1184782	0	.7411444	n = 52
	within		1.101624	-.3736764	5.263029	T-bar = 758.308
PDI score difference	overall	25.5202	18.39061	0	89	N = 27057
	between		7.938871	15.49405	47.59596	n = 38
	within		17.04152	-16.30448	77.07304	T-bar = 712.026
IDV score difference	overall	21.04077	15.66508	0	70	N = 27057
	between		6.58814	13.55952	36.3937	n = 38
	within		14.39696	-12.16332	58.4787	T-bar = 712.026
MAS score difference	overall	27.12311	18.86127	0	95	N = 27057
	between		8.833206	19.28883	59.48485	n = 38
	within		17.39246	-20.36174	84.40787	T-bar = 712.026
UAI score difference	overall	26.08903	18.63695	0	77	N = 27057
	between		6.654332	18.19328	48.23779	n = 38
	within		17.18218	-16.14875	72.64585	T-bar = 712.026

(Continued)

LTO score difference	overall	20.41897	14.12771	0	69	N = 27057
	between		6.557217	14.44503	39.16667	n = 38
	within		12.93736	-11.17287	58.57344	T-bar = 712.026
IND score difference	overall	21.71889	15.71548	0	65	N = 25122
	between		5.189289	11.05643	31.90625	n = 37
	within		14.98135	-10.18736	55.56063	T-bar = 678.973
Official voting system	overall	2.169507	.7965558	1	3	N = 39432
	between		.4503864	1	3	n = 52
	within		.7242571	.6627502	3.424409	T-bar = 758.308
Country voting system	overall	2.134611	.8181434	1	3	N = 39432
	between		.4396539	1	3	n = 52
	within		.7556907	.3346115	3.513696	T-bar = 758.308
Deviation official	overall	.0439998	.2050972	0	1	N = 39432
voting system	between		.1051708	0	.4412533	n = 52
	within		.189611	-.3972535	1.025773	T-bar = 758.308
Order of appearance	overall	11.88623	6.834452	1	28	N = 39432
	between		.3943922	10.27778	12.37455	n = 52
	within		6.82973	.5116891	28.21397	T-bar = 758.308
Host country	overall	.0302039	.1711501	0	1	N = 39432
	between		.0071214	.0207006	.0555556	n = 52
	within		.1710781	-.0253517	1.009503	T-bar = 758.308
Reputation	overall	.7563907	1.223516	0	6	N = 39432
	between		.1311504	.2222222	.9891892	n = 52
	within		1.220175	-.2327984	6.252005	T-bar = 758.308
English song	overall	.5524954	.4972429	0	1	N = 39432
	between		.1502105	.1111111	.827027	n = 52
	within		.4848551	-.2745316	1.441384	T-bar = 758.308
French song	overall	.0622084	.2415367	0	1	N = 39432
	between		.0329782	.027027	.2222222	n = 52
	within		.2406791	-.1600139	1.035181	T-bar = 758.308
German song	overall	.0303307	.1714978	0	1	N = 39432
	between		.0265854	0	.1111111	n = 52
	within		.1703929	-.0807804	1.028426	T-bar = 758.308
Male singer	overall	.448349	.4973321	0	1	N = 35343
	between		.0244901	.4239008	.5714286	n = 52
	within		.4970552	-.1230795	1.024448	T-bar = 679.673
Duet	overall	.1114304	.3146685	0	1	N = 38625
	between		.0080599	.0934783	.132626	n = 52
	within		.3145975	-.0211956	1.017952	T-bar = 742.788
Group	overall	.2164142	.4118052	0	1	N = 38625
	between		.0355902	.1388889	.3888889	n = 52
	within		.4114395	-.1724746	1.077525	T-bar = 742.788

(Continued)

Reciprocity	overall	3.049484	3.782825	0	12	N = 9296
	between		1.054288	1.153846	5.166276	n = 34
	within		3.643863	-2.116793	13.59975	T-bar = 273.412
Lagged vote	overall	3.049484	3.782825	0	12	N = 9296
	between		.1823458	2.388889	3.625	n = 34
	within		3.781917	-.5755164	12.21139	T-bar = 273.412

Appendix B

Model 1:

Random-effects GLS regression — Number of obs = 25,122
Group variable: **CountryID** — Number of groups = 37

R-sq:
within = 0.0632
between = 0.2495
overall = 0.0639

Obs per group:
min = 18
avg = 679.0
max = 1,019

Wald chi2(36) = .
corr(u_i, X) = 0 (assumed) — Prob > chi2 = .

(Std. Err. adjusted for 37 clusters in CountryID)

Voting_bias	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
Shared_border	.4319413	.1376989	3.14	0.002	.1620564	.7018262
Shared_language	-.1988476	.137749	-1.44	0.149	-.4688306	.0711354
Shared_religion	-.0680428	.0845619	-0.80	0.421	-.233781	.0976955
Minority_score	.2517398	.0290525	8.66	0.000	.1947979	.3086816
PDI_diff	-.00421	.0013164	-3.20	0.001	-.00679	-.0016299
IDV_diff	-.0011211	.001644	-0.68	0.495	-.0043433	.002101
MAS_diff	-.0030307	.0010405	-2.91	0.004	-.00507	-.0009913
UAI_diff	.0003807	.0013992	0.27	0.786	-.0023617	.003123
LTO_diff	-.0030424	.0015629	-1.95	0.052	-.0061056	.0000208
IND_diff	.0009638	.0019105	0.50	0.614	-.0027806	.0047083
_cons	2.671863	.1630201	16.39	0.000	2.352349	2.991376
sigma_u	0					
sigma_e	1.9151553					
rho	0	(fraction of variance due to u_i)				

Note: year fixed effects have been included but are omitted from presentation.

Model 2:

Random-effects GLS regression — Number of obs = 22,568
Group variable: **CountryID** — Number of groups = 37

R-sq:
within = 0.0764
between = 0.2195
overall = 0.0766

Obs per group:
min = 14
avg = 609.9
max = 910

Wald chi2(36) = .
corr(u_i, X) = 0 (assumed) — Prob > chi2 = .

(Std. Err. adjusted for 37 clusters in CountryID)

Voting_bias	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
Shared_border	.4282124	.1439949	2.97	0.003	.1459876	.7104372
Shared_language	-.1935235	.1220819	-1.59	0.113	-.4327997	.0457527
Shared_religion	-.2108412	.0978896	-2.15	0.031	-.4027013	-.0189811
Minority_score	.2714612	.0309842	8.76	0.000	.2107332	.3321892
PDI_diff	-.0038686	.0014283	-2.71	0.007	-.0066681	-.0010691
IDV_diff	-.0001979	.0017103	-0.12	0.908	-.00355	.0031542
MAS_diff	-.0028884	.0009215	-3.13	0.002	-.0046945	-.0010823
UAI_diff	-.0006988	.0013457	-0.52	0.604	-.0033363	.0019387
LTO_diff	-.0037125	.0016376	-2.27	0.023	-.0069221	-.0005029
IND_diff	.0007162	.0019372	0.37	0.712	-.0030808	.0045131
Order_of_appearance	.0194004	.0015994	12.13	0.000	.0162656	.0225353
Host_country	.3280484	.0608751	5.39	0.000	.2087354	.4473613
Reputation	.007988	.0144883	0.55	0.581	-.0204086	.0363846
Number_of_voters	-.019259	.0014508	-13.27	0.000	-.0221025	-.0164155
English_song	.1486431	.0550798	2.70	0.007	.0406886	.2565976
French_song	-.0253056	.0669015	-0.38	0.705	-.1564302	.1058189
German_song	-.0751232	.0659444	-1.14	0.255	-.204372	.0541255
Male_singer	-.0071193	.0280423	-0.25	0.800	-.0620811	.0478425
Duet	.1731753	.0533638	3.25	0.001	.068584	.2777665
Group	-.0916008	.0386787	-2.37	0.018	-.1674095	-.015792
_cons	2.820434	.1767719	15.96	0.000	2.473968	3.166901
sigma_u	0					
sigma_e	1.8947472					
rho	0	(fraction of variance due to u_i)				

Note: year fixed effects have been included but are omitted from presentation.

Model 3:

Random-effects GLS regression — Number of obs = 6,366
Group variable: **CountryID** — Number of groups = 29

R-sq:
within = 0.0594
between = 0.1707
overall = 0.0604

Obs per group:
min = 12
avg = 219.5
max = 382

Wald chi2(28) = .
corr(u_i, X) = 0 (assumed) — Prob > chi2 = .

(Std. Err. adjusted for 29 clusters in CountryID)

Voting_bias	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
Shared_border	.2026252	.1255706	1.61	0.107	-.0434887	.4487391
Shared_language	-.3543622	.121853	-2.91	0.004	-.5931897	-.1155347
Shared_religion	-.0611498	.0583822	-1.05	0.295	-.1755768	.0532772
Minority_score	.1311667	.0375713	3.49	0.000	.0575284	.204805
PDI_diff	-.001861	.0022071	-0.84	0.399	-.0061868	.0024648
IDV_diff	-.0026153	.0019583	-1.34	0.182	-.0064534	.0012229
MAS_diff	-.0049218	.0018271	-2.69	0.007	-.0085027	-.0013408
UAI_diff	-.000372	.002532	-0.15	0.883	-.0053347	.0045907
LTO_diff	-.0076346	.0024527	-3.11	0.002	-.0124418	-.0028275
IND_diff	.0017001	.0027927	0.61	0.543	-.0037734	.0071737
Order_of_appearance	.0242792	.0043432	5.59	0.000	.0157668	.0327916
Host_country	.3650594	.1120826	3.26	0.001	.1453815	.5847372
Reputation	.060444	.0250902	2.41	0.016	.0112681	.1096199
Number_of_voters	0	(omitted)				
English_song	.4949553	.0969099	5.11	0.000	.3050154	.6848952
French_song	.3548441	.082598	4.30	0.000	.192955	.5167332
German_song	.2255258	.0774829	2.91	0.004	.0736621	.3773896
Male_singer	-.2928259	.0440401	-6.65	0.000	-.3791429	-.2065089
Duet	.0350826	.1005066	0.35	0.727	-.1619068	.232072
Group	-.0130835	.0474944	-0.28	0.783	-.1061707	.0800038
Reciprocity	.005453	.0069751	0.78	0.434	-.0082179	.019124
Lagged_vote	.0251623	.0073738	3.41	0.001	.01071	.0396146
_cons	2.122683	.1574228	13.48	0.000	1.81414	2.431226
sigma_u	0					
sigma_e	2.0030311					
rho	0	(fraction of variance due to u_i)				

Note: year fixed effects have been included but are omitted from presentation.

Model 4:

```
Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Random-effects tobit regression                 Number of obs     =      6,366
                                                  Uncensored      =      6,169
Limits: lower = 0                                 Left-censored   =        196
        upper = 12                                Right-censored  =          1

Group variable: CountryID                       Number of groups  =         29
Random effects u_i ~ Gaussian                   Obs per group:
                                                              min =         12
                                                              avg =      219.5
                                                              max =        382

Integration method: mvaghermite                 Integration pts.  =         12

                                                Wald chi2(48)     =   1.60e+06
Log likelihood  = -13376.526                    Prob > chi2       =     0.0000
```

(Replications based on **29** clusters in CountryID)

Voting_bias	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
Shared_border	.2138021	.1203352	1.78	0.076	-.0220507	.4496548
Shared_language	-.3460997	.1540323	-2.25	0.025	-.6479974	-.0442019
Shared_religion	-.0504491	.0795697	-0.63	0.526	-.2064029	.1055047
Minority_score	.1331779	.0377069	3.53	0.000	.0592736	.2070821
PDI_diff	-.002415	.0025586	-0.94	0.345	-.0074298	.0025997
IDV_diff	-.0024048	.002175	-1.11	0.269	-.0066676	.0018581
MAS_diff	-.0049794	.0021701	-2.29	0.022	-.0092327	-.0007261
UAI_diff	.0000923	.002732	0.03	0.973	-.0052623	.005447
LTO_diff	-.007673	.0025182	-3.05	0.002	-.0126086	-.0027373
IND_diff	.0022279	.0032113	0.69	0.488	-.0040661	.008522
Order_of_appearance	.0262917	.0044155	5.95	0.000	.0176376	.0349459
Host_country	.3973937	.0899329	4.42	0.000	.2211285	.5736588
Reputation	.0556247	.0210669	2.64	0.008	.0143344	.096915
Number_of_voters	0	(omitted)				
English_song	.5164093	.0914978	5.64	0.000	.3370768	.6957418
French_song	.394688	.082921	4.76	0.000	.2321658	.5572101
German_song	.1665722	.0791775	2.10	0.035	.0113871	.3217573
Male_singer	-.335804	.047953	-7.00	0.000	-.4297901	-.2418178
Duet	.0762858	.0821732	0.93	0.353	-.0847706	.2373422
Group	.0251462	.0482919	0.52	0.603	-.0695043	.1197966
Reciprocity	.0046383	.0076975	0.60	0.547	-.0104485	.0197251
Lagged_vote	.0266504	.0077226	3.45	0.001	.0115145	.0417864
_cons	2.060277	.1622538	12.70	0.000	1.742265	2.378288
/sigma_u	.0412862	.0508923	0.81	0.417	-.0584609	.1410333
/sigma_e	2.046077	.0286238	71.48	0.000	1.989975	2.102178
rho	.000407	.0010012			1.40e-06	.0221246

LR test of sigma u=0: chibar2(01) = **0.04** Prob >= chibar2 = **0.421**

Note: year fixed effects have been included but are omitted from presentation.

Appendix C

Model 5:

Random-effects GLS regression — Number of obs = 22,568
Group variable: **CountryID** — Number of groups = 37

R-sq:
within = 0.0789
between = 0.2241
overall = 0.0795

Obs per group:
min = 14
avg = 609.9
max = 910

Wald chi2(36) = .
corr(u_i, X) = 0 (assumed) — Prob > chi2 = .

(Std. Err. adjusted for 37 clusters in CountryID)

Voting_bias	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
Country_voting_system						
2	-.7026246	.1400612	-5.02	0.000	-.9771396	-.4281096
3	-.3635036	.1454539	-2.50	0.012	-.6485879	-.0784192
Shared_border	.4298151	.1448621	2.97	0.003	.1458905	.7137397
Shared_language	-.1791949	.1233243	-1.45	0.146	-.4209062	.0625164
Shared_religion	-.1732246	.1108157	-1.56	0.118	-.3904194	.0439702
Minority_score	.2720755	.0309823	8.78	0.000	.2113512	.3327997
PDI_diff	-.004424	.0014486	-3.05	0.002	-.0072632	-.0015847
IDV_diff	-.0004069	.0016974	-0.24	0.811	-.0037337	.00292
MAS_diff	-.0029492	.0008921	-3.31	0.001	-.0046977	-.0012006
UAI_diff	-.0001569	.0012375	-0.13	0.899	-.0025823	.0022685
LTO_diff	-.0036707	.00164	-2.24	0.025	-.0068851	-.0004564
IND_diff	.0006383	.0019396	0.33	0.742	-.0031632	.0044398
Order_of_appearance	.0194379	.0015972	12.17	0.000	.0163075	.0225683
Host_country	.3281518	.0604918	5.42	0.000	.20959	.4467136
Reputation	.0059774	.0144558	0.41	0.679	-.0223554	.0343103
Number_of_voters	-.0207632	.0014395	-14.42	0.000	-.0235846	-.0179417
English_song	.1471107	.0548896	2.68	0.007	.039529	.2546924
French_song	-.0300484	.0668794	-0.45	0.653	-.1611297	.1010329
German_song	-.0782259	.0659688	-1.19	0.236	-.2075224	.0510705
Male_singer	-.0037387	.028314	-0.13	0.895	-.0592331	.0517557
Duet	.1724662	.0536557	3.21	0.001	.0673029	.2776295
Group	-.0876049	.0386974	-2.26	0.024	-.1634504	-.0117593
_cons	2.811646	.174698	16.09	0.000	2.469244	3.154047
sigma_u	0					
sigma_e	1.892122					
rho	0	(fraction of variance due to u_i)				

Note: year fixed effects have been included but are omitted from presentation.

Model 6:

Random-effects GLS regression — Number of obs = 22,568
Group variable: **CountryID** — Number of groups = 37

R-sq:
within = 0.0792
between = 0.2276
overall = 0.0798

Obs per group:
min = 14
avg = 609.9
max = 910

Wald chi2(36) = .
corr(u_i, X) = 0 (assumed) — Prob > chi2 = .

(Std. Err. adjusted for 37 clusters in CountryID)

Voting_bias	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
Country_voting_system						
2	-.5156027	.1164945	-4.43	0.000	-.7439278	-.2872776
3	-.2447553	.1177119	-2.08	0.038	-.4754664	-.0140442
dev_voting_system	.2714342	.0942766	2.88	0.004	.0866555	.4562129
Shared_border	.4295862	.1451484	2.96	0.003	.1451006	.7140719
Shared_language	-.1801047	.1233204	-1.46	0.144	-.4218082	.0615989
Shared_religion	-.172626	.1113705	-1.55	0.121	-.3909083	.0456562
Minority_score	.2721115	.0309661	8.79	0.000	.2114191	.332804
PDI_diff	-.0044783	.0014475	-3.09	0.002	-.0073153	-.0016414
IDV_diff	-.0004608	.0016875	-0.27	0.785	-.0037682	.0028465
MAS_diff	-.0029911	.0008896	-3.36	0.001	-.0047347	-.0012474
UAI_diff	-.000092	.0012306	-0.07	0.940	-.0025039	.00232
LTO_diff	-.0036527	.0016394	-2.23	0.026	-.0068659	-.0004395
IND_diff	.0006463	.0019463	0.33	0.740	-.0031683	.0044609
Order_of_appearance	.0194225	.0015969	12.16	0.000	.0162926	.0225525
Host_country	.3283259	.0604757	5.43	0.000	.2097957	.4468561
Reputation	.0056482	.0144297	0.39	0.695	-.0226335	.03393
Number_of_voters	-.0204218	.0014306	-14.27	0.000	-.0232258	-.0176179
English_song	.1469089	.0548468	2.68	0.007	.0394112	.2544066
French_song	-.0302348	.0667885	-0.45	0.651	-.1611379	.1006682
German_song	-.0768172	.0660373	-1.16	0.245	-.206248	.0526136
Male_singer	-.0045929	.0283105	-0.16	0.871	-.0600805	.0508947
Duet	.1726572	.053662	3.22	0.001	.0674817	.2778328
Group	-.088149	.0386016	-2.28	0.022	-.1638068	-.0124912
_cons	2.805836	.174657	16.06	0.000	2.463515	3.148158
sigma_u	0					
sigma_e	1.8918855					
rho	0	(fraction of variance due to u_i)				

Note: year fixed effects have been included but are omitted from presentation.

Model 7:

```
Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Random-effects tobit regression                 Number of obs     =     22,568
                                                  Uncensored      =     22,191
Limits: lower = 0                                 Left-censored   =        376
        upper = 12                                Right-censored  =          1

Group variable: CountryID                       Number of groups  =         37
Random effects u_i ~ Gaussian                   Obs per group:
                                                              min =         14
                                                              avg =      609.9
                                                              max =        910

Integration method: mvaghermite                 Integration pts.  =         12

                                                Wald chi2(49)     =          .
Log likelihood  = -46324.473                    Prob > chi2       =          .

                                  (Replications based on 37 clusters in CountryID)
```

Voting_bias	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
Country_voting_system						
2	-.5426155	.1975724	-2.75	0.006	-.9298502	-.1553808
3	-.2146115	.1613847	-1.33	0.184	-.5309196	.1016967
dev_voting_system	.2499074	.1621761	1.54	0.123	-.0679518	.5677667
Shared_border	.4611855	.156498	2.95	0.003	.154455	.767916
Shared_language	-.2050057	.1334179	-1.54	0.124	-.4664999	.0564885
Shared_religion	-.2268302	.1288252	-1.76	0.078	-.479323	.0256626
Minority_score	.2705707	.039671	6.82	0.000	.1928171	.3483243
PDI_diff	-.0039585	.00166	-2.38	0.017	-.0072121	-.000705
IDV_diff	.0005026	.0015749	0.32	0.750	-.002584	.0035893
MAS_diff	-.0032697	.001011	-3.23	0.001	-.0052513	-.0012882
UAI_diff	-.0008698	.0015874	-0.55	0.584	-.003981	.0022414
LTO_diff	-.0042413	.0018854	-2.25	0.024	-.0079366	-.0005459
IND_diff	.0003849	.0020929	0.18	0.854	-.0037172	.004487
Order_of_appearance	.020259	.0019172	10.57	0.000	.0165014	.0240166
Host_country	.306232	.0556575	5.50	0.000	.1971452	.4153187
Reputation	.0118717	.0167092	0.71	0.477	-.0208777	.0446211
Number_of_voters	-.0209416	.0015714	-13.33	0.000	-.0240215	-.0178617
English_song	.1564959	.0579701	2.70	0.007	.0428765	.2701152
French_song	-.0041578	.0719231	-0.06	0.954	-.1451244	.1368088
German_song	-.1203145	.0744837	-1.62	0.106	-.2662999	.0256708
Male_singer	-.0185027	.0328886	-0.56	0.574	-.0829633	.0459578
Duet	.1905216	.0462883	4.12	0.000	.0997983	.2812449
Group	-.0929835	.0396917	-2.34	0.019	-.1707778	-.0151892
_cons	2.863919	.2462058	11.63	0.000	2.381365	3.346474
/sigma_u	.1304932	.0306127	4.26	0.000	.0704933	.190493
/sigma_e	1.914178	.0238239	80.35	0.000	1.867483	1.960872
rho	.0046259	.0021459			.0017822	.0109857

```
LR test of sigma_u=0: chibar2(01) = 41.31            Prob >= chibar2 = 0.000
```

Note: year fixed effects have been included but are omitted from presentation.

Appendix D

Model 8:

Random-effects GLS regression — Number of obs = 22,568
Group variable: **CountryID** — Number of groups = 37

R-sq:
within = 0.0969
between = 0.1222
overall = 0.0974

Obs per group:
min = 14
avg = 609.9
max = 910

Wald chi2(36) = .
corr(u_i, X) = 0 (assumed) — Prob > chi2 = .

(Std. Err. adjusted for 37 clusters in CountryID)

Voting_bias	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
Country_voting_system						
2	-.3839637	.3164563	-1.21	0.225	-1.004207	.2362793
3	-.4107392	.2386935	-1.72	0.085	-.8785698	.0570914
dev_voting_system	.317121	.0964222	3.29	0.001	.128137	.506105
Order_of_appearance	.018638	.0017067	10.92	0.000	.0152928	.0219831
Host_country	.28333	.0613574	4.62	0.000	.1630718	.4035883
Reputation	.0111637	.0144406	0.77	0.439	-.0171394	.0394669
Number_of_voters	-.0201001	.0014534	-13.83	0.000	-.0229487	-.0172514
English_song	.1478133	.0540822	2.73	0.006	.0418142	.2538125
French_song	-.0547483	.0715418	-0.77	0.444	-.1949677	.0854711
German_song	-.0105795	.0786708	-0.13	0.893	-.1647714	.1436123
Male_singer	-.003622	.0278719	-0.13	0.897	-.0582499	.0510059
Duet	.182304	.0547379	3.33	0.001	.0750196	.2895884
Group	-.0783029	.0358526	-2.18	0.029	-.1485727	-.008033
1.Shared_border	.1197241	.1202166	1.00	0.319	-.1158961	.3553442
Shared_border#Country_voting_system						
1 2	.4318124	.339677	1.27	0.204	-.2339423	1.097567
1 3	.3966815	.2398811	1.65	0.098	-.0734769	.8668399
1.Shared_language	-.1340416	.114211	-1.17	0.241	-.357891	.0898078
Shared_language#Country_voting_system						
1 2	.0277581	.2551647	0.11	0.913	-.4723555	.5278717
1 3	.1011274	.2210417	0.46	0.647	-.3321064	.5343612
1.Shared_religion	.1819185	.111446	1.63	0.103	-.0365116	.4003487
Shared_religion#Country_voting_system						
1 2	-.6872033	.3037494	-2.26	0.024	-1.282541	-.0918654
1 3	-.4879562	.2705864	-1.80	0.071	-1.018296	.0423834
Minority_score	.0498668	.0279193	1.79	0.074	-.0048541	.1045877
Country_voting_system#c.Minority_score						
2	.4770032	.084937	5.62	0.000	.3105298	.6434766
3	.1981099	.0614842	3.22	0.001	.0776031	.3186166
PDI_diff	-.0085719	.002031	-4.22	0.000	-.0125526	-.0045911
Country_voting_system#c.PDI_diff						
2	.0051156	.0038874	1.32	0.188	-.0025036	.0127348
3	.0038879	.0027588	1.41	0.159	-.0015192	.009295
IDV_diff	-.0060054	.0031778	-1.89	0.059	-.0122338	.000223
Country_voting_system#c.IDV_diff						
2	.0086732	.0053257	1.63	0.103	-.001765	.0191115
3	.0072171	.0036698	1.97	0.049	.0000244	.0144098
MAS_diff	-.0052984	.0017885	-2.96	0.003	-.0088037	-.0017931
Country_voting_system#c.MAS_diff						
2	.0009565	.0036112	0.26	0.791	-.0061213	.0080343
3	.0038168	.0026304	1.45	0.147	-.0013387	.0089723
UAI_diff	.003675	.0027619	1.33	0.183	-.0017383	.0090882
Country_voting_system#c.UAI_diff						
2	-.0057202	.0037942	-1.51	0.132	-.0131568	.0017163
3	-.0049881	.0031041	-1.61	0.108	-.0110721	.0010959
LTO_diff	-.0080817	.0025437	-3.18	0.001	-.0130673	-.0030961
Country_voting_system#c.LTO_diff						
2	.0026512	.0035299	0.75	0.453	-.0042672	.0095696
3	.0085084	.0029101	2.92	0.003	.0028046	.0142121
IND_diff	.0006374	.0026831	0.24	0.812	-.0046215	.0058962
Country_voting_system#c.IND_diff						
2	.0004181	.0041281	0.10	0.919	-.0076728	.008509
3	.002443	.0032062	0.76	0.446	-.003841	.008727
_cons	2.863097	.2573303	11.13	0.000	2.358739	3.367455
sigma_u	0					
sigma_e	1.8743913					
rho	0	(fraction of variance due to u_i)				

Note: year fixed effects have been included but are omitted from presentation.

Average Marginal effects, based on Model 8:

```
Average marginal effects                        Number of obs     =     22,568
Model VCE    : Robust

Expression   : Linear prediction, predict()
dy/dx w.r.t. : 1.Shared_border 1.Shared_language 1.Shared_religion Minority_score PDI_diff IDV_diff MAS_diff UAI_di
               LTO_diff IND_diff

1._at        : Country_voting_system=          1

2._at        : Country_voting_system=          2

3._at        : Country_voting_system=          3
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf. Interval]	
0.Shared_border	(base outcome)					
1.Shared_border						
_at						
1	.1197241	.1202166	1.00	0.319	-.1158961	.3553442
2	.5515364	.3053521	1.81	0.071	-.0469426	1.150015
3	.5164056	.1884647	2.74	0.006	.1470216	.8857895
0.Shared_language	(base outcome)					
1.Shared_language						
_at						
1	-.1340416	.114211	-1.17	0.241	-.357891	.0898078
2	-.1062835	.2393522	-0.44	0.657	-.5754051	.3628381
3	-.0329143	.1915924	-0.17	0.864	-.4084284	.3425999
0.Shared_religion	(base outcome)					
1.Shared_religion						
_at						
1	.1819185	.111446	1.63	0.103	-.0365116	.4003487
2	-.5052847	.2261486	-2.23	0.025	-.9485278	-.0620417
3	-.3060376	.1878076	-1.63	0.103	-.6741337	.0620584
Minority_score						
_at						
1	.0498668	.0279193	1.79	0.074	-.0048541	.1045877
2	.52687	.0752506	7.00	0.000	.3793816	.6743584
3	.2479767	.0444619	5.58	0.000	.160833	.3351204
PDI_diff						
_at						
1	-.0085719	.002031	-4.22	0.000	-.0125526	-.0045911
2	-.0034563	.0029059	-1.19	0.234	-.0091518	.0022392
3	-.004684	.0017096	-2.74	0.006	-.0080348	-.0013332

(Continued on the next page)

IDV_diff						
_at						
1	-.0060054	.0031778	-1.89	0.059	-.0122338	.000223
2	.0026678	.0034484	0.77	0.439	-.0040909	.0094266
3	.0012117	.0020086	0.60	0.546	-.002725	.0051484
MAS_diff						
_at						
1	-.0052984	.0017885	-2.96	0.003	-.0088037	-.0017931
2	-.0043419	.0023893	-1.82	0.069	-.0090247	.000341
3	-.0014816	.0012455	-1.19	0.234	-.0039228	.0009596
UAI_diff						
_at						
1	.003675	.0027619	1.33	0.183	-.0017383	.0090882
2	-.0020453	.0022443	-0.91	0.362	-.006444	.0023534
3	-.0013132	.0012469	-1.05	0.292	-.0037569	.0011306
LTO_diff						
_at						
1	-.0080817	.0025437	-3.18	0.001	-.0130673	-.0030961
2	-.0054305	.0027385	-1.98	0.047	-.0107978	-.0000632
3	.0004266	.001818	0.23	0.814	-.0031366	.0039898
IND_diff						
_at						
1	.0006374	.0026831	0.24	0.812	-.0046215	.0058962
2	.0010554	.0029888	0.35	0.724	-.0048025	.0069134
3	.0030804	.0020271	1.52	0.129	-.0008926	.0070534

Note: dy/dx for factor levels is the discrete change from the base level.