LG-STV-PR – Voter Preferences Rankings – Research and Analysis

Scottish Local Government Elections May 2022

Des McNelis

10/09/24

Abbreviations
PR = Proportional Representation
STV = Single Transferable Vote
WIGM = Weighted Inclusive Gregory Method
LGA = Local Government Authority
LG-STV-PR elections = Local Government STV-PR
RO = Returning Officer
BLT = Preference Profile files (file extension often used is .BLT – BaLloT)
GP = Geometric Progression
H/W = Hardware
S/W = Software

Introduction

This paper concerns itself with the particular Proportional Representation (PR) method adopted for use in the Scottish Local Government Elections in May 2022. This method used the Single Transferable Vote (STV) with the WIGM variant (Weighted Inclusive Gregory Method). The aim of STV-PR is to offer fairer voter representation, in particular avoiding "wasted" votes.
Note that this PR method is not that used for electing MSPs to the Scottish Parliament.

A brief introduction to STV is here.
The keywords to take away are quota, transfers, preferences.

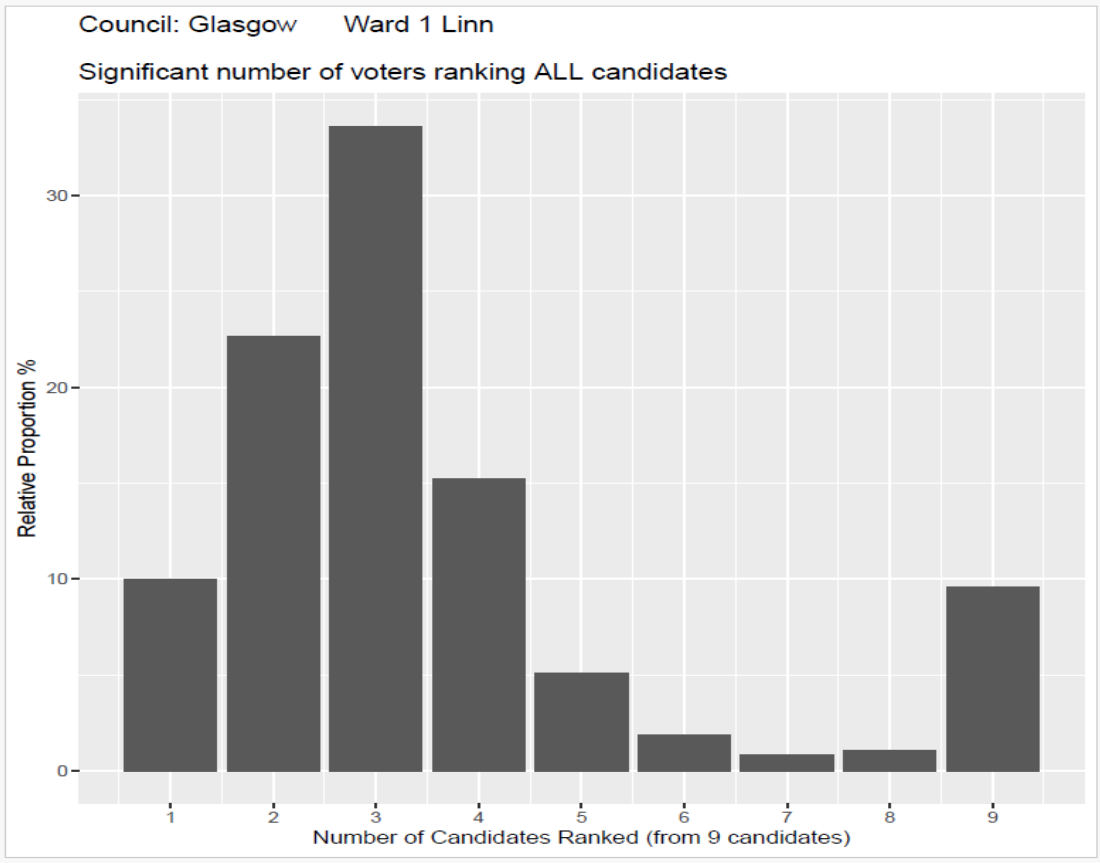
The associated ballot paper style requires the voter to rank their candidate preferences by placing a "1" alongside the name of their first preferred candidate. If the voter so desires, they can then place a "2" alongside the name of their 2nd preferred candidate, and so on.

My own experience of completing the ballot paper on election day led me to consider the difficulties my fellow voters may experience. It is difficult enough ranking up to 10 or so choices given time, comfort and due consideration. However, with the minute or so most voters take to complete the ballot paper – in an unusual environment – with some time pressure – I find it difficult to accept that most voters give due consideration to the ranking. Of course, some voters may have given considerable prior thought to any intermediate ranking – but I think they are in the minority. I am not convinced that the typical voter has the necessary sophistication and/or knowledge to successfully "game" the system.
Around 8% of the voters specified the maximum number of preferences – I will refer to this voter behaviour as "**deep preferences ranking**".

I believe that there is significant misunderstanding by political parties and voters in this area – and an over-confidence in the merit of deep preferences ranking. My analysis – over the 348 contested wards – confirms that ranking more than five or six preferences has little or no impact on candidate election outcomes.

The following graph is typical. The graph appears to consist of two parts – the first part is akin to a statistical Normal Distribution curve – and the second part appears unexpected (to me). What is the explanation of the high proportion of voters ranking ALL candidates?
Here are my thoughts:

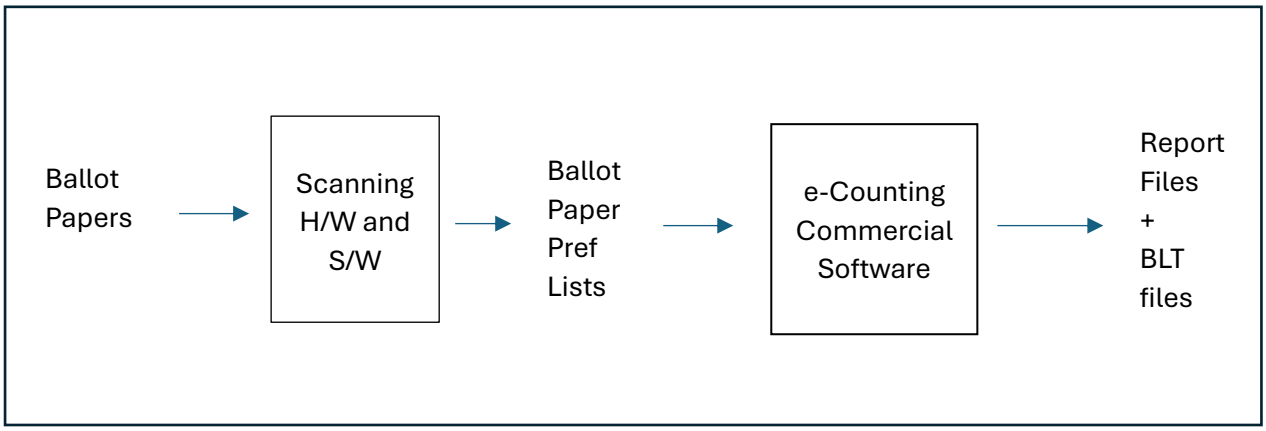
- Many voters appear to think they have to apply preferences for many/all candidates.
- Do some voters – or groups of voters – believe in some kind of "anyone but ...." mechanism whereby ranking their least favourite candidate last results in the minimum credit – some kind of "tactical voting"?



All the graphs for Glasgow can be viewed in the attached document Preferences_Graphs__Glasgow__All_wards.pdf. It can be seen that most wards exhibit this behaviour.

The following diagram describes the journey ballot papers take through the Election e-Counting System. The BLT file structure is documented later.

Diagram 1: Election e-Counting System flow.



The BLT file.

The BLT file is a basic text file, although it is internally formatted in a particular style. It is generated by the e-Counting Software after the scanning of the ballot papers and the STV-PR computation has been completed.

There is one BLT file produced for each ward. The BLT file contains information on how voters ranked the candidates for the ward.

The BLT file contains *all* the data sufficient to (re)build the following election reports:

- Candidate Votes per Stage Report
- Transfers Report

The BLT file is of particular import to the following stakeholders:

- Political party analysts.
- Academic researchers and analysts.

There follows a full description of the data contained within a BLT file

Sample Preference Profile (BLT) file: Council: Glasgow. Ward: 01 Linn.

```
9 4                                  Number of candidates.      Number of seats.
5 6 2 4 8 0                          5 preference patterns of   6 2 4 8
3 4 6 9 3 8 2 5 7 1 0
1 2 7 1 0                            Trailing zero signifies the end
1 8 2 1 5 7 9 0                      of a preference pattern.
..............
..............
1 2 8 1 5 9 4 6 3 7 0
4 4 6 3 9 7 5 8 2 1 0
1 6 7 9 4 0
0                                    Signifies end of preferences lists.
"Euan BLOCKLEY" "Scottish Conservative and Unionist"
"Malcolm CUNNING" "Glasgow Labour"
"Angela JONES" "Alba Party for independence"                Candidate name
"Paul MCCABE" "Scottish National Party (SNP)"               and party.
"Joe MCCAULEY" "Scottish Liberal Democrats"
"Margaret MORGAN" "Scottish National Party (SNP)"
"James TONER" ""
"Catherine VALLIS" "Glasgow Labour"
"Keith WARWICK" "Scottish Greens - Delivering For Our Community"
"Ward 1 Linn"                                               Ward name.
```

The first preference pattern indicates that five voters ranked the candidates in the same way:

- First preference: Candidate 6 (Margaret MORGAN)
- Second preference: Candidate 2 (Malcolm CUNNING)
- Third preference: Candidate 4 (Paul MCCABE)
- Fourth preference: Candidate 8 (Catherine VALLIS)
- No further preference declared.

There is some evidence of voters completing preferences “down the page” or “alphabetically”:

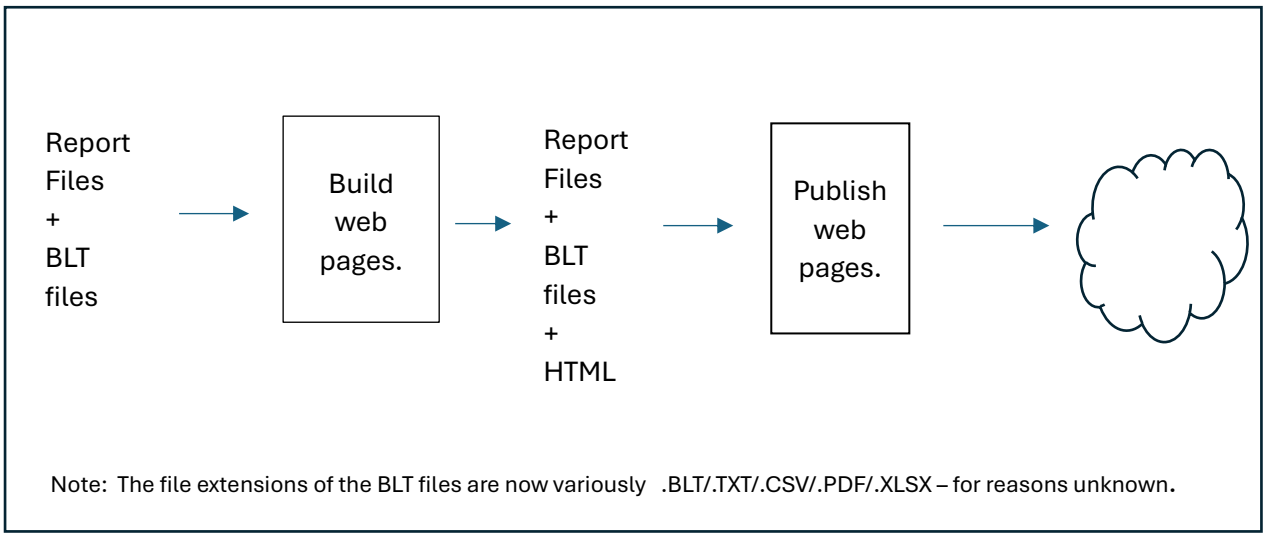
```
3 1 2 3 4 5 6 7 8 9 10 0
1 5 1 2 3 4 6 7 8 9 10 0
3 1 2 3 5 4 6 7 8 9 10 0
```

(Glasgow – Ward 02).

Discovered *en passant* – there may be many more similar occurrences. This is another area worthy of further analysis (this is a well-known area of research – but with little consensus on a sensible solution).

The following diagram describes how the files output by the Election e-Counting software are ultimately published on each of the 32 councils' websites.

Diagram 2: Publication of Report and BLT files on 32 Council websites.



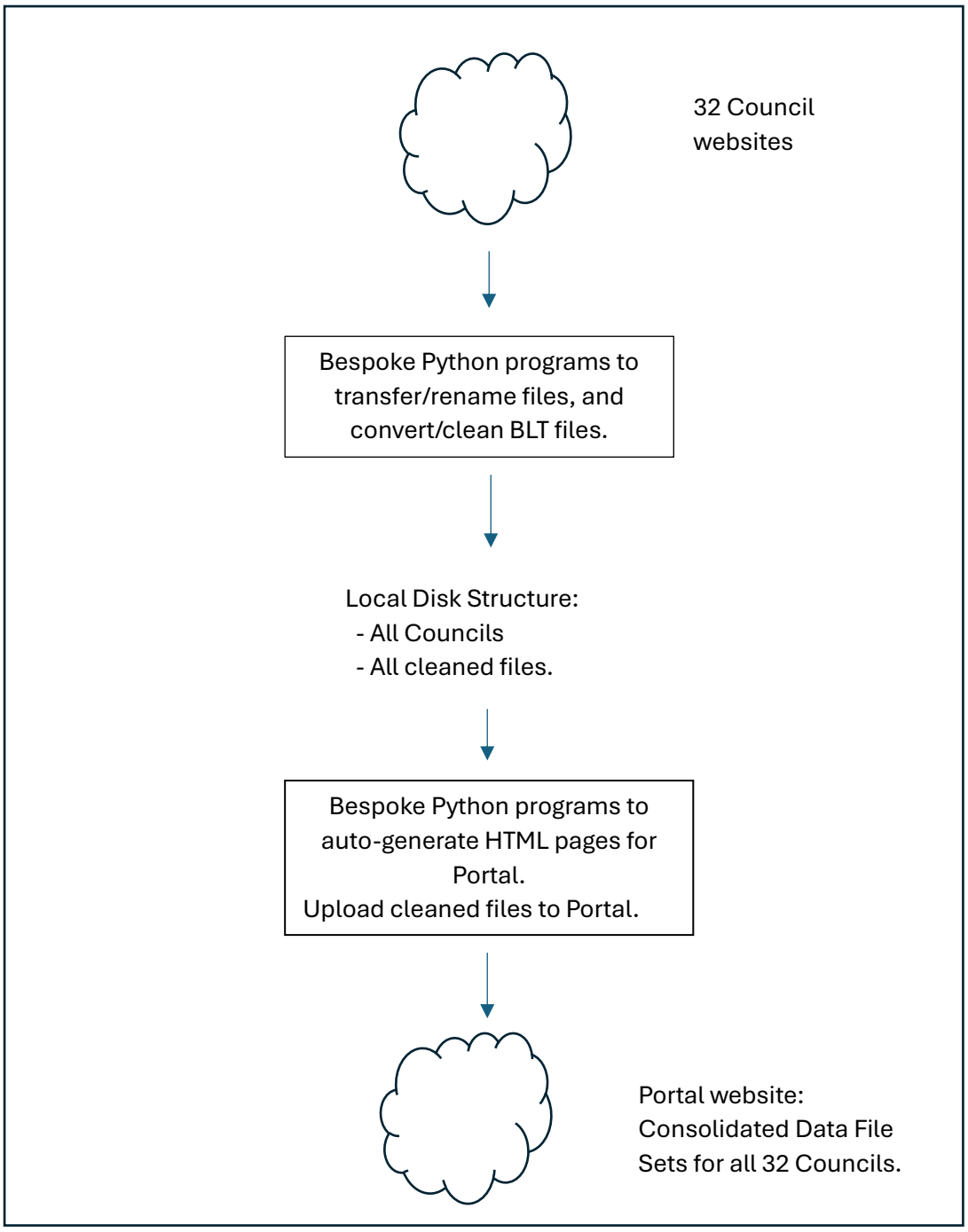
The data files for the 32 Scottish Councils can be viewed by visiting each of the 32 websites individually.
Here are the 32 direct hyperlinks:
https://stv.prorep.org.uk/introduction/the-32-scottish-councils

I have downloaded all the available report and BLT files from all of the 355 wards. These files were then cleaned-up (both filenames and content) and then used to build a dedicated website which then allows a consistent and easy method of viewing and downloading the files (also in ZIP containers at both national and council level).

Consolidated Election Results Portal: https://stv-results.prorep.org.uk

The following diagram illustrates this process.

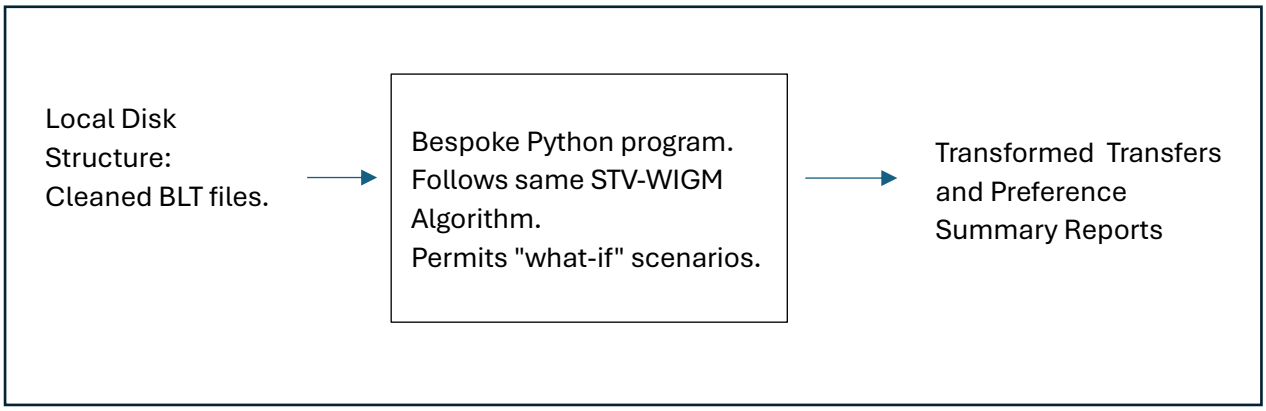
Diagram 3: Build Portal website: Consolidated Data File Sets for all Councils.



I have written a Python program which emulates the same STV-WIGM Algorithm used in the Scottish Local Government Elections May 2022.

This Local Disk Structure is then used to re-run the Scottish Local Government Elections May 2022 – but now with different input parameters to enable various “what-if” scenarios. Only the BLT files are required.

Diagram 4: Bespoke Python program to re-run Scottish Local Government Elections May 2022 with different parameters.



We now have the capability to re-run the Scottish Local Government Elections with a varying number of preferences. We can then compare the report outputs of the parameterised election re-run with the original reports to establish matching results. A match occurs when the same set of candidates is elected for a ward – re-run versus original.

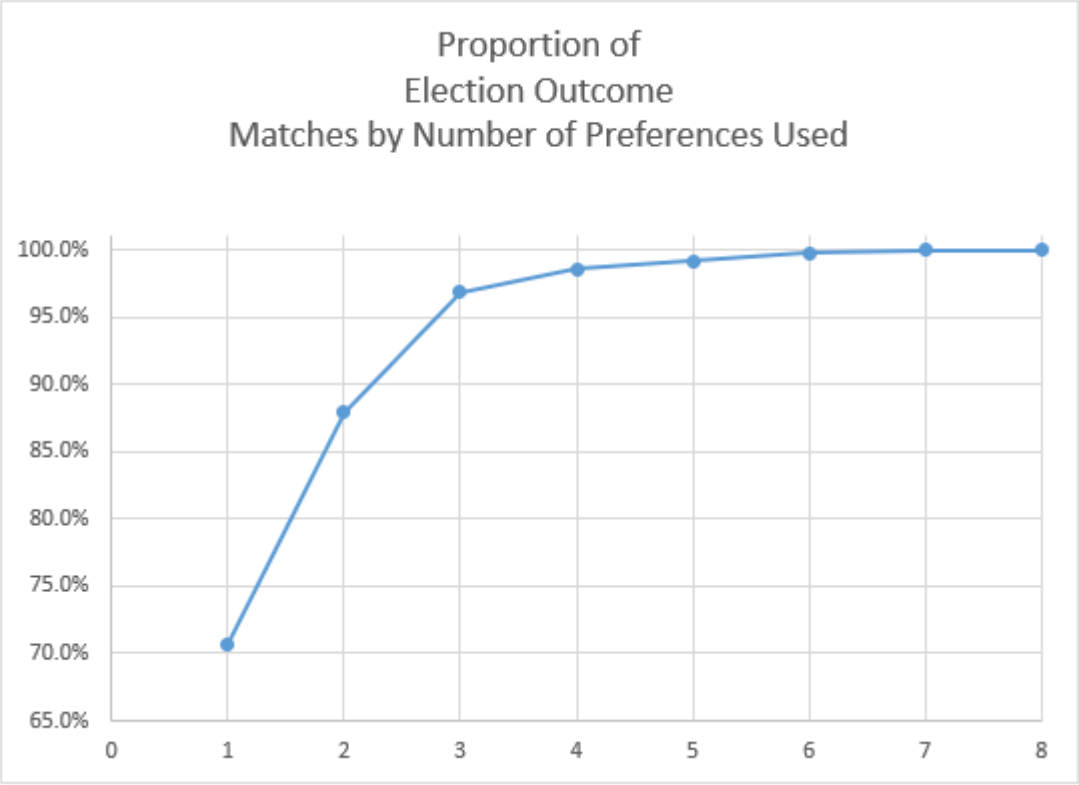
The following table lists those matches against the number of preferences.

Table 1: ALL 348 Contested Wards with matching results (from full reports) for varying number of candidates ranked:

Number of Election Outcome Matches by Number of Preferences Used

Data source: Scottish Local Government Elections May 2022
Denominator = **348 Contested Wards** (355 total wards - 7 Uncontested wards)

Number of Preferences Used	Number of Matches	Number of Matches %
1	246	70.7%
2	306	87.9%
3	337	96.8%
4	343	98.6%
5	345	99.1%
6	347	99.7%
7	348	100.0%
8	348	100.0%



Example: The 4th line down in the early table: 4 343 98.6%.
The Python program has re-run the election – but now only considering the first 4 preferences.
There were only five mismatches (348 - 343) – and so the matching rate was 343/348 = 98.6%.

There were no mismatches beyond considering only the first 6 preferences. So, for this election, no preferences beyond the 6th had any impact on the overall election outcomes.

Costs and downsides of deep preferences ranking.

If indeed there are issues here regarding ranking many candidates, then there are associated costs and downsides:

- Voters may be disinclined to vote at all because of the perceived complexity in completing the ballot paper.
- Voters may make incorrect choices while completing the ballot paper which rather invalidates the entire voting process.
- Voters may overthink the voting process.
- Those voters who attempt to vote tactically – or simply in error – inadvertently promote those other candidates in the middle (or end) of the preference profile pattern. That's clearly unfair.
- The ballot paper scanning process is more complex.
- Far more ballot papers have to be passed to the Adjudicators and RO – this wastes time and may result in misclassified ballot paper images. See Appendix 1 for more information.

Appendix 2 reinforces my stance. Modelling the STV process as a Geometric Progression (GP) produces similar results as regards "deep preferences ranking".

Conclusions

I believe that there is significant misunderstanding by parties and voters in this area – and an over-confidence in the merit of deep preferences ranking. My analysis – over the 348 contested wards – confirms that ranking more than five or six preferences has little or no impact on candidate election outcomes.

References

The Scottish Local Government Elections Order 2007 – The LG STV-PR election "rules"
https://www.legislation.gov.uk/ssi/2007/42/contents/made
especially ....
https://www.legislation.gov.uk/ssi/2007/42/schedule/1/made

Consolidated Election Results Portal (All Councils): https://stv-results.prorep.org.uk

How does the Single Transferable Vote work in council elections?
https://news.stv.tv/politics/how-does-the-single-transferable-vote-work-in-scottish-council-elections

Acknowledgements

I would like to thank Dr James Gilmour for his advice and support over the last 12 months or so.

Attachments

- Spreadsheet: Election Re-run Matches - Preferences.xlsx
- Spreadsheet: Election Re-run Matches - Geometric Progression.xlsx
- Preferences_Graphs__Glasgow__All_wards.pdf
- Sample output: Transfers Report - My text style - Glasgow.txt . View this file with a fixed-pitch font like Courier New. Or open with a basic text editor like Notepad or Notepad++.

Appendix 1 Very High Ballot Paper Scanning Error Rate.

The document "Scottish Local Government Elections 2022 Electronic Vote Counting Factsheet" https://www.emb.scot/downloads/file/934/ecounting-2022-factsheet-local-government-elections-2022-ecounting-overview
has a paragraph:

"If a ballot paper does not pass the IT system checks, the scanned image is sent to an Adjudicator. This person views the ballot paper image to see if the voter's intentions can be confirmed. In many cases, the voter's intentions will be obvious when the ballot paper is viewed. A common question is why the ballot paper needs reviewed when it looks clear. This is often because, for example, what might look like a "1" to a person might also look a bit like a "7" to an IT system. The IT system is set up to send anything doubtful to an Adjudicator**. Typically, between 15% and 20% of ballot papers need to be reviewed by an Adjudicator**." (my emboldening).

This appears to me to be a very high ballot paper scanning/classification error rate. The classification of handwritten digits is the 101 class of Machine Learning (ML) – it is the "Hello World" of ML. I would expect the correct classification rate to be much higher than 80%-85% – nearer 95%.

The MNIST Database of handwritten digits:
https://en.wikipedia.org/wiki/MNIST_database#:~:text=The%20MNIST%20database%20(Modified%20National,the%20field%20of%20machine%20learning.

Regardless, *unnecessary* deep preferences ranking further complicates the scanning process.

Appendix 2 Model LG-STV-PR as a Geometric Progression (GP)

Further analysis I have completed is modelling the STV calculation process as a mathematical geometric progression https://www.mathsisfun.com/algebra/sequences-sums-geometric.html

I used only the relative proportion of preferences (akin to the Preference Summary Report). It is important to note that I made no use whatsoever of vote transfers. This has proved most interesting: A GP common ratio of 0.30 resulted in a "hit ratio" of 78%.

Glasgow City Council

Preference Summary Report

This report details all candidates preferences over count stages.

Contest Name	Ward 1 Linn
Total Number of Valid Votes	8,994
Quota	1,799
Percentage Turnout	41.2

Candidate Name	Affiliation	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
Euan BLOCKLEY	Scottish Conservative and Unionist	1,035	319	318	156	71	37	54	125	504
Malcolm CUNNING	Glasgow Labour	2,152	1,387	720	364	255	157	126	94	6
Angela JONES	Alba Party for independence	114	120	301	323	140	147	183	227	83
Paul MCCABE	Scottish National Party (SNP)	1,934	1,256	652	206	112	69	77	75	45
Joe MCCAULEY	Scottish Liberal Democrats	572	616	797	444	218	238	163	137	25
Margaret MORGAN	Scottish National Party (SNP)	1,070	2,032	617	225	107	78	74	74	53
James TONER		849	332	506	337	244	126	137	125	83
Catherine VALLIS	Glasgow Labour	724	1,634	846	546	288	191	128	47	18
Keith WARWICK	Scottish Greens - Delivering For Our Community	544	399	1,302	434	229	163	95	56	48
	Totals	8,994	8,095	6,059	3,035	1,664	1,206	1,037	960	865

Example: Glasgow Ward 1 Linn
The data used is that derived to produce the Preference Summary Report (example above).
For each candidate, I calculated a total "vote" value as if the 1st/2nd/3rd///...preferences followed a GP with the given common ratio.
So, for Euan BLOCKLEY, for a common ratio of 0.30, the calculation was:

$$\text{Total} = 1{,}035 + 0.3 \times 319 + 0.3^2 \times 318 + 0.3^3 \times 156 + 0.3^4 \times 71 + \ldots.$$

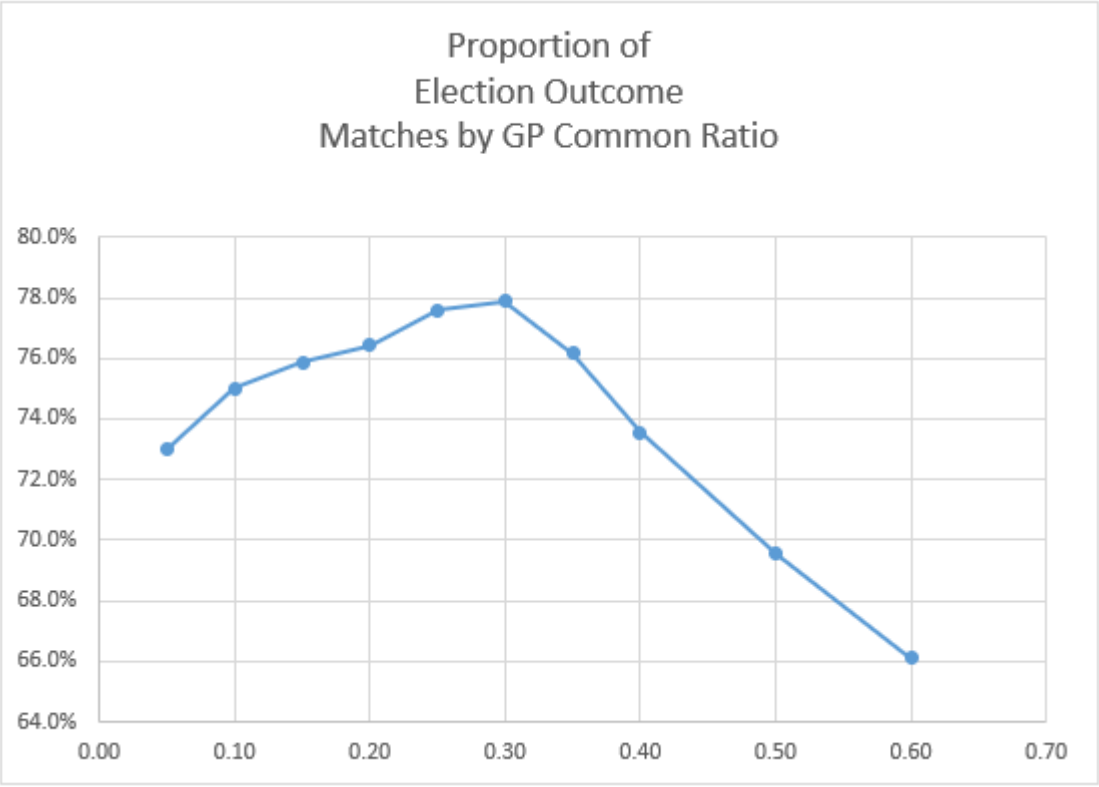
In the following spreadsheet image, the "Contribution of a 5th Preference Vote" and "... 10th Preference Vote" numbers indicate the relative value of one vote if indeed this was a GP with a Common Ratio of 0.30.

Model LG STV-PR as a Geometric Progression

Number of Election Outcome Matches by Geometric Progression Common Ratio

Data source: Scottish Local Government Elections May 2022
Denominator = **348 Contested Wards** (355 total wards - 7 Uncontested wards)

GP Common Ratio	Number of Matches	Number of Matches %	Contribution of a 5th Preference Vote	Contribution of a 10th Preference Vote
0.05	254	73.0%	0.000006	0.000000000002
0.10	261	75.0%	0.000100	0.000000001000
0.15	264	75.9%	0.000506	0.000000038443
0.20	266	76.4%	0.001600	0.000000512000
0.25	270	77.6%	0.003906	0.000003814697
0.30	271	77.9%	0.008100	0.000019683000
0.35	265	76.1%	0.015006	0.000078815639
0.40	256	73.6%	0.025600	0.000262144000
0.50	242	69.5%	0.062500	0.001953125000
0.60	230	66.1%	0.129600	0.010077696000



A GP with a Common Ratio of 0.30 provides a good fit to a full STV-PR-WIGM election calculation. This hints that there is powerful effect of diminishing returns.

Appendix 3 Possible future research.

STV-PR-WIGM – Number of decimal places used in Calculation.
For the current set of rules, 5 d.p. are used for the display of transfers. I can now re-run the election with reducing numbers of d.p. to confirm when there is a difference in outcomes.

"Down-the-page" or "alphabetical" voting.
By analysing the BLT files I may be able to indicate more precisely the frequency of this issue.

District Magnitude – 3-seat and 4-seat wards.
The STV-PR-WIGM algorithm improves in "fairness" as the number of seats per ward increases. I may be able to research this further.

Electorate per seat.
It seems unfair that some seats (mainly rural) require fewer votes to attain than some other seats (mainly urban). I can, initially, analyse the votes-per-seat ratios.

--- End ---