Does the urban-rural divide shape public confidence towards the federal government in Malaysia?

Introduction

The project explores the question, "H*ow does the urban-rural divide shape public confidence towards the (federal) Malaysian government?*". The topic is particularly relevant given the results of recent elections in June 2018 that brought to an end over five decades of rule by the Barisan Nasional government. Initial statistical analyses of the results suggests a swing of the rural vote, previously regarded as a 'safe deposit' for the government, had swung slightly to the opposition and other more conservative parties such as Parti Islam Se-Malaysia. Against this backdrop, the essay wishes to analyse the factors that affect public opinion towards the federal government in Malaysia and whether there indeed exists a rural-urban cleavage where this is concerned. This narrative of a conservative, government-aligned rural positioned against more wealthy and educated populations based in suburban and urban areas is not new and persisted even up until the previous elections in 2013 (Ufen, 2013). Indeed, UMNO, a constituent party of the ruling Barisan Nasional coalition won in 78 out of the 99 rural Malay constituencies, but only in 5 of the 20 urban Malay seats; in Kuala Lumpur, for instance, the BN won only 12.3 per cent support[1].

Consequently, the independent variable that I am concerned with is the built environment that a voter resides in, whether that is rural, suburban or urban. The dependent variable that I will examine is the confidence that voters have in the central government in Malaysia. To do so, I will utilise cross-sectional data from the sixth wave of the World Values Survey. Interestingly, the survey for Malaysia was carried out in 2012, slightly before the 2013 Malaysian General elections and after the Wall Street Journal began reporting about potential corrupt wrongdoings at state-backed investment fund 1Malaysia Development Berhad (1MDB).

Data and Methodology

To explore the relationship between place and preferences, I will use responses from the World Values Survey Wave 6 from 2012. The WVS dataset is a survey-based dataset that is collated by WVS association and its partner institutions across the world in 100 countries. The sample from each country, which is governed by certain guidelines, has to be representative of all people in the age 18 and older residing within private households in each country, regardless of their nationality, citizenship or language. The WVS recommends its affiliates to that a number of respondents per a PSU (or a route in case of random route sample) is not exceeding 10 respondents. It is possible to have several Primary Sampling Units per one settlement; they should be located in quite a good distance from each other. The main method of collection is face-to-face interview at a respondent's place of residence. This is either done through a paper questionnaire or a computer-assisted interview. For Malaysia, Orient Pacific Century carried out the survey under the direction of Prof. Eduard J. Bonhoff and Dr. Grace Lee Hooi Yean. In this case, a proportional to size (PPS) sampling technique was used. All households in sampling frame (N=1,300) were divided into 14 states. This was further stratified based on the type of settlement (urban/rural) based on the urban-rural proportions at the state level. This was then further stratified by race, gender and age.

For the purpose of this research, I have utilised variable *V115* which asks the question, "I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?: The government (in your nation's capital)", as the dependent variable of confidence in the government. Responses were thereafter coded on an ordinal scale where 1 indicated a great deal of confidence and 4 none at all. Our key predictor variable of concern is that of the geographical origin of a respondent which in this case is represented via the variable

[1] Andreas Ufen. "The 2013 Malaysian Elections: Business as Usual or Part of a Protracted Transition?" *Journal of Current Southeast Asia Affairs,* no. 32, 3-17. https://journals.sub.uni-hamburg.de/giga/jsaa, p. 13.

V253 which records the size of the town in which the response is collected from. This is an ordinal variable that ranges from 1 (indicating a town of under 2000 residents) to 8 (indicating a town with over 500 000 residents). Other variables I will explain in greater detail in the following sections include controls for income, education level, importance of politics to respondent's life, national pride, political orientation, trust in persons, ethnicity, confidence in economy, acceptance of corruption, perception of democratic governance and age.

Prior to analysis, I cleaned the data by removing all values below 0 (which are not useful) and recoded some.

Dependent variable

[V115, Confidence in government] The dependent variable of concern is the degree of confidence that a respondent has in the government. In the initial coding, WVS assigned a value of 1 to a response indicating 'a great deal of confidence'. For ease of interpretation, I have inverted the coding with 1 indicating a score of 'no confidence at all' to 4, indicating strong confidence. In this case, the WVS asks respondents about their sentiment with regards to several institutions with this variable clearly alluding to the political leadership of the day, separate from the civil service, judiciary and central bank. This variable is referred to as *conf_gov* in our models.

Key independent variable

[V253, Size of town] The independent variable of concern is the locale in which the respondent is from. The WVS coding uses an ordinal scale to represent the continuous variable of persons in a town. In this case, the scoring system is as follows. This is the *townsize* variable in the models.

Size of town

1	Under 2,000
2	2,000-5,000
3	5,000-10,000
4	10,000-20,000
5	20,000-50,000
6	50,000-100,000
7	100,000-500,000
8	500,000 and more

I have recoded this into several forms, namely, (i) as a nominal variable (*areatype*) with 3 categories (rural-suburban-urban), (ii) ordinal variable (*townsizecat)* with collapsed levels (Under 50 000 as '0', over 50000 as '1'. The models will consider all three types of the same variable. The rationale for altering this is that there exists an official classification of area type based on population size, meaning that the space that one is from could be captured by agglomerating the original ordinal variables under clear categories.

Control variables

1. **[V239, Perceived income strata]** Existing political science literature draws clear relationships between one's income or economic standing in society and voting patterns (Huber and Stanig, 2009). As such, it is wise to consider this variable in context of understanding a respondent's response to confidence in political leadership of the government. To do so, I used responses to the question, "Scales of income", which placed respondents in an ordinal scale from '1' indicating the lower step to '10' the highest step. This is the *income* variable in our model.

2. **[V248, Educational attainment]** Likewise, literature also suggests a strong relationship between education level of a voter and their voting preferences. As such, we

consider responses to the question, "Highest educational level attained", where 1 indicates a situation where no formal education has been attained while 9 indicates university-level education. This variable is coded as *educ_lvl* in the models.

I also consider a recoded version that collapses the levels into 3 ordinal categories where 1 indicates persons who have less than primary education, 2 as a person who has at most high school qualifications and 3 indicating a minimum of some kind of exposure to tertiary education. This is referred to as *educ_lvlrecoded* in our model.

3. **[V84, Importance of politics]** There may be a case to be made for the interest and attention one places to political developments and their corresponding approval of the government. To code this we look at responses to V84 of the WVS survey which asks respondents to rank their interest in politics where 1 indicates a 'very interested' response and 4 as 'not at all interested'. For the ease of interpretation, I recoded the ordinal variable where 1 indicates a complete lack of interest and 4 as the most interested. This is referred to as *import_pol* in our model.

4. **[V95, Position on political spectrum]** One may also imagine that one's position on the political spectrum affecting their approval of the government. For instance, a Republican may approve of a Republican President more than he would a Democratic one. To control for this, I look at responses to the question "self-positioning in political scale" where 1 indicates the most left and 10 the most right. For the ease of interpretation, I reordered this with 1 indicating extreme conservatism and 10 as extreme liberalism. This is referred to as *pol_pol* in our model.

5. **[V24, Life philosophy on trust]** One's intrinsic disposition to trust may also affect their confidence in the government. To control for this I look at responses to V24 which asks for responses to the following statement, "most people can be trusted" where a score of 1 indicates that most people can be trusted and 2 that one needs to be very careful. This is the variable *trust* in our model.

6. **[V254, Ethnicity]** Race-based politics is a topic of concern in the political science literature on Malaysia and one should control for the effect of ethnicity on trust in government. To do so, I coded ethnicity as a factor variable, *ethnicity*, in the models that I will consider.

7. **[V146, Religiosity]** Religion is also a topic of concern in the political science literature. There is concern that the recent Islamisation of the politics have resulted in a fundamental realignment of the political system in the country. However, the research is concerned more with the level of religiosity that one exhibits and the effects that that has on the respondent's confidence in the government. To do so, I look at responses to the question, 'How often do you pray'. A score of 1 denotes 'several times a day ' and 8 as 'practically never'. To ease interpretation, I recoded this where 1 indicates low religiosity and 8 as highest religiosity. This is the *relig* variable in the models.

8. **[I_NATIONALISM, National pride]** The level of national pride may also affect one's perception of the functioning of the government. To measure this, I look at scores on national pride where 0 is very low and 1 is very high. This is coded as *nat_pride* in our model.

9. **[V60, Confidence in economy]** Personal confidence in the economic outlook of the country may also affect one's view of government performance. This is measured using responses to the question, "Aims of country", or their preferred choice of country goal where 1 indicated a high level of economic growth. This aspiration is recoded with ifelse

in R. All other responses were coded with a 0 while economic growth was coded as 1. This is coded as *con_econ* in our model.

10. **[V202, Acceptability of corruption in society]** Corruption, particularly on a lower scale, is present in Malaysia. As such, one's opinion on the acceptability of corruption may also affect how they view government performance, particularly if the government is one that is besieged by corruption allegations. In this case, we take responses to the question "Justifiable: Someone accepting a bribe in the course of their duties" where 1 indicated never justifiable and 10, always justifiable'. This is represented by the *corr_acp* variable in our model.

11. **[V141, Democratic governance]** Perceptions regarding the democratic governance of the country may also affect government confidence. In this case, we consider responses to V141, a question that asks "How democratically is this country being governed today?". Here, 1 indicates a position of not at all democratic and 10, completely democratic. This is represented by the *dem_gov* variable in our model.

12. **[V242, Age]** The last variable that is considered is age which is a continuous variable. We consider the continuous one as well as an ordinal one with 4 levels, Generation Y (16-29), Generation X (30-50), Baby boomers (51-70) and 71-110 (Elderly).

Summary of Data

Table 1: Descriptive statistics for independent, dependent and control variables

Variable	Obs Used	Min	Max	Median	Mean	Std Dev	Kurtosis	Skewness
Income	1299	1	10	6	6	1.84	-0.003397	-0.657329
Conf_gov	**1300**	**1**	**4**	**3**	**2.90**	**0.77**	**0.7790786**	**-0.3329627**
Townsize	**411**	**1**	**8**	**6**	**5.51**	**1.65**	**-0.248903**	**-0.671738**
Townsizecat	**411**	**0**	**1**	**1**	**0.76**	**0.60**	**-1.23886**	**-0.4663438**
Educ_lvl	1300	1	9	5	5.05	1.91	-0.179246	0.1499981
Educ_lvlrecode	1300	1	3	2	1.92	0.57	0.0564951	-0.014655
Import_pol	1300	1	4	3	2.60	0.80	-0.435787	-0.113738
Nat_pride	1300	0	1	0	0.13	0.18	1.132792	1.223812
Pol_pol	1300	1	10	6	6.60	1.92	-0.006205	0.1283948
Trust	1300	1	2	2	1.91	0.28	6.789989	-2.96391
Relig	1300	1	11	8	6.95	1.92	2.016205	-1.82102
Con_econ	1300	0	1	1	0.67	0.47	-1.468744	-0.729646
Corr_acp	1300	1	10	1	2.36	2.24	1.847288	1.658933
Age	1300	18	80	41	40.01	13.96	-0.552177	0.3079813
Dem_gov	1300	1	10	7	7.17	2.24	0.862588	-0.7104504

For our dependent variable of concern, *conf_gov*, the median is greater than the mean indicating negatively skewed sample distribution. In this case, the average respondent replies with a score of 2.90 (indicating a score close to *Quite A Lot*). However, the variance of the score is surprisingly large as indicated by a standard deviation of 0.77. Overall, the sample is relatively unskewed and shows very slight signs of a platykurtic distribution.

For our independent variable of concern, *townsize*, there are 411 usable observations in the sample having removed negative values. The median is once again larger than the mean, indicating a (very slightly) negatively skewed sample distribution. The average town size is 5.51 (indicating an area with a population somewhere between 50,000 to 100,000). The standard deviation is somewhat large at 1.65. Overall, the sample is slightly skewed and shows slight signs of a leptokurtic distribution. Our recoded independent variable of 2 levels has a slightly stronger leptokurtic distribution but slightly less negative skew.

Two control variables are worth discussing in greater detail, *import_pol* and *religion*. *Import_pol* suggests that the average respondent is rather concerned about politics in their country with a mean of around 2.60 though there exists, interestingly, a large standard deviation of around 0.80. This suggests that there are substantial numbers who care a lot and care very little. At the same time, the average person is also rather religious, with a mean of 6.95, indicating prayer on an almost daily basis.

Hypothesis

The hypothesis I wish to test is as follows:

H_0: The origin of a respondent has no effect on the respondent's confidence in the government of the day.
H_1: The origin of a respondent has an effect on the respondent's confidence in the government of the day.

I am postulating that the null hypothesis of no effect will be rejected by a statistically significant coefficient on the t*ownsize* variables. I am of the belief that the more urban the area, the more negative the confidence in government.

Initial Models

I first investigate the relationship between locality size and confidence in government before proceeding to experiment with various multiple linear regression models. The simple regression models I have chosen to discuss as part of the initial exploration are as follows:

$Confidence\ in\ government = \beta_0 + \beta_1 townsize_i + \varepsilon$……(1.1)

$Confidence\ in\ government = \beta_0 + \beta_1 townsizecat_i + \varepsilon$……(1.2)

$Confidence\ in\ government = \beta_0 + \beta_1 Rural_i + \beta_2 Suburban_i + \beta_3 Urban_i + \varepsilon$……(1.3)

Running them, we obtain the following statistical output:

Table 2: Regression output for first three models (simple linear)

Regression Results

	Dependent variable:		
	Confidence in Government		
	Model1.1	Model 1.2	Model 1.3
	(1)	(2)	(3)
Constant	2.701***	2.763***	2.868***
	(2.429, 2.973)	(2.603, 2.923)	(2.652, 3.084)
townsize	0.012		
	(-0.035, 0.060)		
townsizecat		0.008	
		(-0.175, 0.191)	
areatypeSuburban			-0.183
			(-0.435, 0.068)
areatypeUrban			-0.064
			(-0.306, 0.178)
Observations	411	411	411
R2	0.001	0.00002	0.007
Adjusted R2	-0.002	-0.002	0.002
Residual Std. Error	0.805 (df = 409)	0.805 (df = 409)	0.803 (df = 408)
F Statistic	0.263 (df = 1; 409)	0.007 (df = 1; 409)	1.422 (df = 2; 408)
Note:			*p<0.1; **p<0.05; ***p<0.01

The initial models, while producing results, perform very poorly in providing meaningful relationships between the two variables. Indeed, both Model 1.1 and Model 1.2 produce an adjusted R-squared that is negative. Model 1.3, the one which includes town size recoded as area types, performs slightly better but still only predicts just 0.2% of total variance in the sample. It is unsurprising given that all three models are likely to suffer from omitted variable bias.

We move on to more complex models that include multiple predictor variables, adding controls that I believe will similarly affect the dependent variable of concern. The models that will be considered are as follows:

$$Confidence\ in\ government=\beta_0 + \beta_1 Income_i + \beta_2 townsize_i + \beta_3 educlvl_i + \beta_4 corr\_acp_i + \beta_5 religion_i + \beta_6 dem\_gov + \beta_7 pol\_pol_i + \beta_8 Age_i + \beta_9 Malay_i + \beta_{10} Chinese_i + \beta_{11} nat\_pride_i + \varepsilon \quad (1.4)$$

$$Confidence\ in\ government=\beta_0 + \beta_1 Income_i + \beta_2 townsize_i + \beta_3 educlvl_i + \beta_4 corr\_acp_i + \beta_5 religion_i + \beta_6 dem\_gov + \beta_7 pol\_pol_i + \beta_8 Age_i + \beta_9 Malay_i + \beta_{10} Chinese_i + \beta_{11} nat\_pride_i + \boldsymbol{\beta_{12} trust_i + \beta_{13} import\_pol_i} + \varepsilon \quad (1.5)$$

$$Confidence\ in\ government=\beta_0 + \beta_1 Income_i + \boldsymbol{\beta_1 Rural_i + \beta_2 Suburban_i + \beta_3 Urban_i} + \beta_4 educlvl_i + \beta_5 corr\_acp_i + \beta_6 religion_i + \beta_7 dem\_gov_i + \beta_8 pol\_pol_i + \beta_9 Age_i + \beta_{10} Malay_i + \beta_{11} Chinese_i + \beta_{12} nat\_pride_i + \beta_{13} trust_i + \beta_{14} import\_pol_i + \varepsilon \quad (1.6)$$

$$Confidence\ in\ government=\beta_0 + \beta_1 Income_i + \boldsymbol{\beta_2 townsizecat_i} + \beta_3 educlvl_i + \beta_4 corr\_acp_i + \beta_5 religion_i + \beta_6 dem\_gov + \beta_7 pol\_pol_i + \beta_8 Age_i + \beta_9 Malay_i + \beta_{10} Chinese_i + \beta_{11} nat\_pride_i + \beta_{12} trust_i + \beta_{13} import\_pol_i + \varepsilon \quad (1.7)$$

Table 3: Regression output for 4 candidate multiple linear regression models

	Dependent variable:			
	Confidence in Government			
	1.4	1.5	1.6	1.7
	(1)	(2)	(3)	(4)
Constant	2.874***	4.217***	4.446***	4.322***
	(2.115, 3.632)	(3.236, 5.197)	(3.476, 5.416)	(3.367, 5.277)
income	-0.015	-0.018	-0.020	-0.016
	(-0.053, 0.023)	(-0.057, 0.020)	(-0.058, 0.018)	(-0.054, 0.023)
townsize	0.002	0.001		
	(-0.043, 0.046)	(-0.043, 0.044)		
areatypeSuburban			-0.235*	
			(-0.473, 0.002)	
areatypeUrban			-0.105	
			(-0.327, 0.118)	
townsizecat				-0.110
				(-0.280, 0.061)
educ_lvl	-0.027	-0.032	-0.038*	-0.031
	(-0.070, 0.016)	(-0.074, 0.010)	(-0.080, 0.004)	(-0.072, 0.011)
corr_acp	0.050**	0.042**	0.031	0.038*
	(0.010, 0.090)	(0.002, 0.081)	(-0.009, 0.071)	(-0.001, 0.077)
relig	0.008	0.001	-0.003	-0.0001
	(-0.034, 0.050)	(-0.041, 0.042)	(-0.044, 0.039)	(-0.042, 0.041)
dem_gov	0.053***	0.056***	0.057***	0.059***
	(0.014, 0.092)	(0.018, 0.095)	(0.019, 0.095)	(0.020, 0.098)
pol_pol	0.053**	0.052**	0.050**	0.052**
	(0.012, 0.094)	(0.011, 0.093)	(0.009, 0.090)	(0.011, 0.092)
age	-0.005*	-0.006**	-0.006**	-0.006**
	(-0.011, 0.001)	(-0.012, -0.0003)	(-0.012, -0.001)	(-0.012, -0.0003)
as.factor(ethnic)295	-0.361*	-0.363*	-0.340*	-0.372*
	(-0.749, 0.027)	(-0.744, 0.018)	(-0.720, 0.040)	(-0.753, 0.008)
as.factor(ethnic)780	-0.365*	-0.392**	-0.378**	-0.416**
	(-0.735, 0.005)	(-0.755, -0.029)	(-0.741, -0.016)	(-0.778, -0.053)
nat_pride	-1.299***	-1.265***	-1.304***	-1.291***
	(-1.696, -0.902)	(-1.657, -0.874)	(-1.693, -0.914)	(-1.682, -0.900)
trust		-0.506***	-0.515***	-0.516***
		(-0.778, -0.234)	(-0.785, -0.244)	(-0.788, -0.244)
import_pol		-0.085*	-0.076	-0.085*
		(-0.177, 0.007)	(-0.168, 0.016)	(-0.176, 0.007)
Observations	411	411	411	411
R2	0.193	0.227	0.236	0.230
Adjusted R2	0.171	0.202	0.209	0.205
Residual Std. Error	0.732 (df = 399)	0.718 (df = 397)	0.715 (df = 396)	0.717 (df = 397)
F Statistic	8.684*** (df = 11; 399)	8.959*** (df = 13; 397)	8.720*** (df = 14; 396)	9.117*** (df = 13; 397)
Note:				*p<0.1; **p<0.05; ***p<0.01

Following the results in the simple regression models 1.1,1.2 and 1.3, I proceeded to run the 4 aforementioned multiple regression models. Overall, they have far better overall adjusted R-squared (on average explaining around 19.68% of the total variance) than the simple regression models. I will first begin by discussing the results of Model 1.5. This model includes all our initial control, independent and dependent variables.

The formula is as follows:

$$Confidence\ in\ government = 4.217 - 0.018Income_i + 0.001townsize_i - 0.032educlvl_i + 0.042corr\_acp_i + 0.001religion_i + 0.056dem\_gov + 0.052pol\_pol_i - 0.006Age_i - 0{,}363Malay_i - 0.392Chinese_i - 1.265nat\_pride_i - 0.506trust_i - 0.085import\_pol_i + \varepsilon \quad (1.5)$$

In this model, income, town size, education level and religion are not statistically significant. Using the *varImp* function[2] in the caret package in R, the three most important variables are: (1) national pride, (2) life philosophy on trust and (3) perception of democratic governance in the country. In the case of national pride, a score turning from 0 to 1 results in a 1.265 decrease in the score ascribed to the confidence a respondent has to the government controlling for all other factors. The more national pride one has, the more someone is likely to be critical of government performance. Second, a change from trusting to not trusting someone results in a 0.506 decrease in the score given to government performance holding all other variables constant. This suggests that persons who are less trusting in general tend to rate government performance more poorly. The third variable, perception about democratic governance, suggests that each point increase in the perception of democratic governance results in a 0.056 increase in the government confidence score holding all other factors constant. All three variables are statistically significant at the 99.9% level with a corresponding p-value of less than 0.001. This model has an adjusted R-squared of 0.202.

In models 1.6 and 1.7 we alter the regression model by adding different forms of the town size variable – one with two levels and the other with three categories. Model 1.6, the one that contains the categorical variables, returned statistically significant results for only the Suburban category (*) and we an overall adjusted R-square of 0.209. Model 1.7 returned a different coefficient but the same non-statistical significance and a slightly higher adjusted R-squared that model 1.5 had (0.205). The *varImp* function, however, suggests that town size variable with an adjusted factor level increases the importance of it by 3847% from 0.03278 to 1.2612. Moreover, it makes intuitive sense for the areas to have collapse factors that more represent the official classifications by Malaysian authorities. We will nonetheless continue to consider all three variants.

However, there is a need to consider interaction variables. In this case, I wanted to observe how town sizes interacts with importance each person placed on politics and whether the combined effect was different on the score that each respondent gave to the confidence in government score. Given that opposition wards are mainly in more populated suburban and urban areas, it may be wise to consider whether importance of politics interact will locality of respondent.

We thus consider another 3 models with interaction between these two variables:

$$Confidence\ in\ government = \beta_0 + \beta_1 Income_i + \beta_2 townsize_i + \beta_3 educc\_lvlrecode_i + \beta_4 corr\_acp_i + \beta_6 dem\_gov + \beta_7 pol\_pol_i + \beta_8 Age_i + \beta_9 Malay_i + \beta_{10} Chinese_{i+} \beta_{11} nat\_pride_i + \beta_{12} trust_i + \beta_{13} import\_pol_i + \beta_{14} townsize * import\_pol_i + \varepsilon \quad (1.8)$$

$$Confidence\ in\ government = \beta_0 + \beta_1 Income_i + \beta_2 townsizecat_i + \beta_3 educc\_lvlrecode_i + \beta_4 corr\_acp_i + \beta_6 dem\_gov + \beta_7 pol\_pol_i + \beta_8 Age_i + \beta_9 Malay_i + \beta_{10} Chinese_{i+} \beta_{11} nat\_pride_i + \beta_{12} trust_i + \beta_{13} import\_pol_i + \beta_{14} townsizecat * import\_pol_i + \varepsilon \quad (1.9)$$

$$Confidence\ in\ government = \beta_0 + \beta_1 Income_i + \beta_2 areatype_i + \beta_3 educc\_lvlrecode_i + \beta_4 corr\_acp_i + \beta_6 dem\_gov + \beta_7 pol\_pol_i + \beta_8 Age_i + \beta_9 Malay_i + \beta_{10} Chinese_{i+} \beta_{11} nat\_pride_i + \beta_{12} trust_i + \beta_{13} import\_pol_i + \beta_{14} areatype * import\_pol_i + \varepsilon \quad (1.10)$$

[2] The varImp function uses recursive partitioning to obtain the reduction in the loss function (e.g. mean squared error) attributed to each variable at each split is tabulated and the sum is returned. It then ranks them based on this loss function.

Table 4: Regression output for 4 candidate multiple linear regression models

	Confidence in Government		
	1.8	1.9	1.10
	(1)	(2)	(3)
Constant	5.422***	4.906***	4.667***
	(4.254, 6.590)	(3.906, 5.906)	(3.598, 5.736)
income	-0.022	-0.011	-0.025
	(-0.060, 0.016)	(-0.048, 0.026)	(-0.063, 0.012)
townsize	-0.215***		
	(-0.353, -0.077)		
townsizecat		-0.735***	
		(-1.289, -0.182)	
areatypeSuburban			-0.116
			(-0.852, 0.621)
areatypeUrban			-0.816**
			(-1.499, -0.133)
educ_lvlrecode	-0.109	-0.098	-0.125*
	(-0.249, 0.031)	(-0.238, 0.042)	(-0.264, 0.015)
corr_acp	0.045**		0.038*
	(0.006, 0.083)		(-0.002, 0.077)
dem_gov	0.054***	0.060***	0.055***
	(0.016, 0.091)	(0.022, 0.099)	(0.018, 0.093)
pol_pol	0.059***	0.059***	0.060***
	(0.019, 0.100)	(0.018, 0.099)	(0.019, 0.100)
age	-0.007**	-0.007**	-0.007**
	(-0.013, -0.001)	(-0.013, -0.001)	(-0.013, -0.001)
as.factor(ethnic)295	-0.318*	-0.350*	-0.278
	(-0.695, 0.059)	(-0.729, 0.029)	(-0.655, 0.098)
as.factor(ethnic)780	-0.339*	-0.412**	-0.331*
	(-0.693, 0.015)	(-0.768, -0.056)	(-0.684, 0.022)
nat_pride	-1.278***	-1.230***	-1.265***
	(-1.663, -0.892)	(-1.616, -0.844)	(-1.650, -0.880)
trust	-0.485***	-0.548***	-0.462***
	(-0.754, -0.217)	(-0.816, -0.280)	(-0.731, -0.193)
import_pol	-0.570***	-0.261***	-0.231*
	(-0.875, -0.265)	(-0.441, -0.081)	(-0.469, 0.007)
townsize:import_pol	0.087***		
	(0.034, 0.139)		
townsizecat:import_pol		0.235**	
		(0.030, 0.440)	
areatypeSuburban:import_pol			-0.023
			(-0.302, 0.256)
areatypeUrban:import_pol			0.292**
			(0.028, 0.555)
Observations	411	411	411
R2	0.248	0.233	0.259
Adjusted R2	0.223	0.210	0.231
Residual Std. Error	0.709 (df = 397)	0.715 (df = 398)	0.705 (df = 395)
F Statistic	10.052*** (df = 13; 397)	10.091*** (df = 12; 398)	9.190*** (df = 15; 395)
Note:			*p<0.1; **p<0.05; ***p<0.01

As suspected the results for the town size and importance of politics interaction was highly statistically significant across all three models. These 3 models removed from consideration religiosity of a respondent. To adjudicate amongst these three models, I utilised the partial F-test and considered Ockham's Razor for model simplicity. In this case, I decided to proceed with Model 1.9 even though it had the lowest R-squared of 0.216 amongst the three. However, the ones with higher R-squared were either (i) not statistically significant in the partial F-test, (ii) were hard to interpret as an interaction model and (iii) had serious collinearity issues not addressable by mean centering. We now proceed to discuss initial results of Model 1.9, test model assumptions and address the problems using further statistical tools.

Statistical checks

Table 4: vif output for Model 1.9

income	**townsizecat**	**educ_lvlrecode**	**dem_gov**	**pol_pol**
1.116204	11.577678	1.347483	1.094594	1.205570
age	**as.factor(ethnic)295**	**as.factor(ethnic)780**	**nat_pride**	**trust**
1.330231	4.632212	4.715723	1.255437	1.021306
import_pol	**townsizecat:import_pol**			
4.520727	14.849008			

Unsurprisingly, the test for collinearity reveals that the two interacted variables exceed the acceptable vif score of under 5. As such, we will seek to address this issue through mean-centering the *import_pol* variable only since *townsizecat* is a binary variable.

***Table 5: Studentized Breusch-Pagan test for Model** 9*

Data: Model 9
BP = 21.037, df = 12, p-value = **0.03136**

The Breusch-Pagan test suggests that homoscedasticity is also an issue. I will return to this in the mean-centered model.

Table 6: vif output for Model 1.9 with mean-centering

income	**townsizecat**	**educ_lvlrecode**	**nat_pride**
1.158968	1.080699	1.324327	1.258563
con_econ	**dem_gov**	**trust**	**pol_pol**
1.301112	1.098579	1.072460	1.325016
center.import_pol	**agecatgenx**	**agecatbabyboom**	**agecatelderly**
4.636049	1.505456	1.728670	1.158080
as.factor(ethnic)295	**as.factor(ethnic) 780**	**townsizecat:center.import_pol**	
4.656965	4.734358	4.653629	

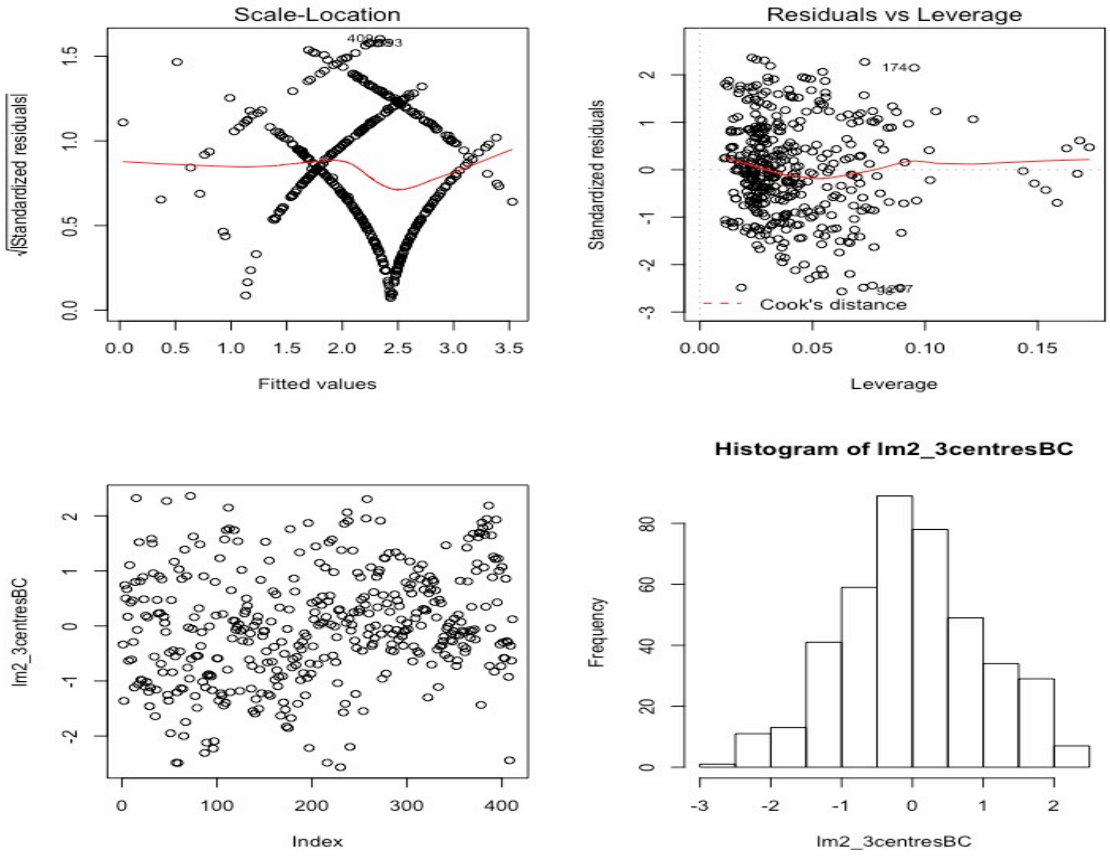
Having mean centered Model 1.9, we now resolve the issue of collinearity but are faced with the issue of homoscedasticity given that it has returned a p-value of <0.05 in the Breusch-Pagan test. To resolve this, we then perform a Box-Cox transformation on the mean centered Model 9. The Box-Cox transformation is valid in this case since we are treating Y as an ordinal variable like a likert scale that is continuous. The lambda provided for the Y from the B-C transformation is 1.3.

Table 7: Studentized Breusch-Pagan test for Model 1.9 with mean-centering

Data: Model 9 centered
BP = 26.342, df = 16, p-value = **0.07186**

We then check whether the residuals of the model are linear, normal and evenly distributed through a visual inspection.

Table 8: Tables for standard residuals for normality, independence, linearity and constant variance assumptions



From the visuals provided above, there appears to be no issue with large outliers, high leverage or influence points fundamentally altering our analysis in the final model.

Final Models and Conclusions

$$Conf\_gov^{1.3} = 4.171 - 0.030Income_i - 0.199townsizecat_i - 0.112educ\_lvlrecode_i + 0.077dem\_gov_i + 0.056pol\_pol_i - 0.198GenX_i - 0.27Babyboom_i - 0.846Elderly_i - 0.404Malay_i - 0.527Chinese_i - 1.557nat\_pride_i - 0.649trust_i - 0.396Center\_import\_pol_i + 0.379townsizecat*Center\_import\_pol_i + E$$

Our final model has an overall adjusted R-squared of 0.217, explaining around 21.7% of the total variance in the sample. It is Model 1.9 with mean-centering and a Box-Cox transformation. The constant in this case has to be edited. We will use a base level of 1 to indicate variables which do not have a 0 as a response. This response will come from someone in an area classified as rural, Malay, with 0 national pride, mean centered importance of politics score of 0 and is from the Gen Y age category. The average score (which has been raised to a power of 1.3) for someone from this constant is 3.109 (2.292 in original scale), indicating a score of 'not very confident'.

Table 9: Regression output for final model

Regression Results

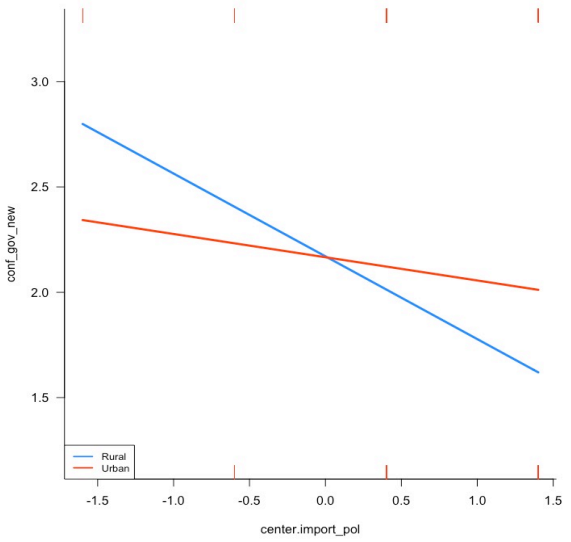
	Conf
	Confidence in Government
	Final Model with B-C transformation
income	-0.030
	(0.026)
townsizecat	-0.199*
	(0.114)
educ_lvlrecode	-0.112
	(0.094)
nat_pride	-1.557***
	(0.260)
con_econ	-0.300**
	(0.122)
dem_gov	0.077***
	(0.026)
trust	-0.649***
	(0.185)
pol_pol	0.056*
	(0.029)
center.import_pol	-0.396***
	(0.123)
agecatgenx	-0.198*
	(0.114)
agecatbabyboom	-0.270*
	(0.154)
agecatelderly	-0.846**
	(0.362)
as.factor(ethnic)295	-0.404
	(0.256)
as.factor(ethnic)780	-0.527**
	(0.240)
townsizecat:center.import_pol	0.379***
	(0.141)
Constant	4.171***
	(0.571)
Observations	411
R2	0.246
Adjusted R2	0.217
Residual Std. Error	0.943 (df = 395)
F Statistic	8.593*** (df = 15; 395)
Note:	*p<0.1; **p<0.05; ***p<0.01

In our final model, the *townsize* categorical variable is not statistically significant but the interaction with the mean-centered importance of politics is very much so. The interpretation is as follows,

(i) Being from an urban area leads to a change in predicted Y lambda, on average, by 0.199 points holding all other factors constant.

(ii) However, the effect from being in an urban area is attenuated by the interaction with the respondent's view of how important politics is. In an urban area, the more involved you are in politics, the more likely you are to be confident of the government as can be seen in the figure below.

Table 10: Interaction effects between townsize categorical and importance of politics



The result leaves my hypothesis in a strange situation. Location does appear to affect confidence in government but only when one considers a respondent's interest in political developments. This is counter-intuitive and the effect of it is also initially hard to grasp. However, on closer inspection, it may be the case that government supporters are likely to pay more attention to politics whilst those in the urban areas who are not confident in the government tend to be more apathetic about politics.

Limitations

There are a few crucial limitations to this study. First are issues related to generalizability. The results are valid only to the sample we are presented with. Although the WVS provides panel data for this, initial explorations suggests that the issue of "Big N, Small T" to be highly problematic. Moreover, given the highly fluid nature of politics, the results may change in the next wave of the WVS. Second, we have made modelling assumptions about constant variance through only the BP test that does not have as strict assumptions as the *ncvTest* in the car package or the *gvlma* test from the *gvlma* package. Under the stricter restrictions, our model would have just violated those assumptions.

Bibliography

Ufen, Andreas. "The 2013 Malaysian Elections: Business as Usual or Part of a Protracted Transition?" *Journal of Current Southeast Asia Affairs*, no. 32, 3-17. https://journals.sub.uni-hamburg.de/giga/jsaa.