Journal of East Asian Studies 15 (2015), 167–198

The 2013 Malaysian Elections: Ethnic Politics or Urban Wave?

Jason Wei Jian Ng, Gary John Rangel, Santha Vaithilingam, and Subramaniam S. Pillay

In this article we examine the electoral impact of urbanization vis-à-vis ethnicity in Malaysia. We employ a robust econometric technique, the fractional response logit model, on data from the recently concluded thirteenth general election. The findings show that there are both an ethnic effect and an urban effect in determining the distribution of parliamentary seats among the political groups. Strong support for the opposition coalition, Pakatan Rakyat, was evident in urban constituencies, while the ruling coalition, Barisan Nasional, continued to enjoy success in rural constituencies. Although Barisan Nasional is still dependent on Bumiputera support, its success is also dependent on non-Bumiputera support from rural constituencies. However, with declining birthrates among the Chinese electorates, this support may not be forthcoming in future elections. We also provide insights for both coalitions to consider in developing strategies for the next election. KEYWORDS: *Malaysia, thirteenth general election, ethnic politics, fractional logit response model, urbanization*

THERE HAVE BEEN THIRTEEN GENERAL ELECTIONS IN MALAYSIA SINCE IT gained independence from the British in 1957. In all thirteen elections, the same coalition, Barisan Nasional (BN)[1] or National Front, has been returned to power. The thirteenth general election (GE13), held on May 5, 2013, was the most fiercely contested election in Malaysia's history. For the first time, there was a degree of uncertainty about BN's ability to retain power. The opposition coalition, Pakatan Rakyat (PR) or People's Alliance, which was formed in 2008, appeared to have become credible enough to launch a genuine threat to BN's stranglehold on power. However, in the end, BN retained power with a simple but comfortable majority by winning 133 of the 222 parliamentary constituencies contested. This was a slight reduction from the 138 seats that it won in the previous election held in 2008. This outcome again denied BN the two-thirds

majority it had enjoyed from independence in 1957 until 2008. By polling only 47.4 percent of the votes cast, it recorded its lowest ever popular support in history.

Peninsular Malaysia was the major battleground where BN won eighty-five seats against PR's eighty.[2] While this result was the same as in the 2008 election, there was, however, a notable change in the distribution of the seats. BN regained some of the Malay-dominated rural seats it had lost in 2008, but also lost a number of urban seats in traditional BN strongholds, especially in the southern part of the peninsula. Given that most Chinese Malaysians live in urban areas, Prime Minister Najib Razak subsequently termed this phenomenon the Chinese Tsunami, purportedly reflecting the huge exodus of Chinese voters from BN to PR (Noh 2014).

However, political analysts have observed that while there was an increase in Chinese support for PR, the electoral outcome also saw a major swing in the urban electorate against BN, causing a further widening of Malaysia's rural-urban rift (Boo 2013). One of the PR leaders, Lim Kit Siang, described this as a "Malaysian and urban tsunami" (*The Star* 2013). This debate of whether the electoral outcome was a result of an urban or a Chinese swing against BN has been inconclusive due to the high correlation between urban parliamentary constituencies and the proportion of Chinese Malaysian voters in any given constituency.

Against this backdrop, the aim of this study is to identify which of the two factors, ethnicity or urbanization, provides a stronger explanation for the erosion of BN's popular votes in GE13. We measure the proportion of votes won by BN due to urbanization and ethnic factors using a robust econometric technique, the fractional response logit model proposed by Papke and Wooldridge (1996). This model extends the generalized linear model by accounting for the bounded nature of the data, which is a proportion quantity.

The article is organized as follows. We begin by providing a description of the parliamentary system and development of politics in Malaysia since its independence. The next section describes the sources of the data and explains the methodology used to model the proportion of votes won by BN, and how the ethnicity and urbanization impacts can be quantified. We then present and discuss the estimation results, concluding with the implications of the results. This article, however, for reasons explained in the next section, only includes election data from Peninsular Malaysia and excludes the two Borneo states of Sabah and Sarawak.

Malaysian Politics and the Electoral System

Historical Background

Peninsular Malaysia, or Malaya as it was then known, became independent from British colonial rule in 1957. In 1963, Singapore and the Borneo states of Sabah and Sarawak merged with Malaya to form a new nation called Malaysia. However, two years later, Singapore separated from Malaysia due to political differences. Since then, Malaysia consists of thirteen states: eleven in Peninsular Malaysia (i.e., the old Malaya) and the two states of Sabah and Sarawak.[3]

In our study we focus on electoral data only from Peninsular Malaysia for several reasons. First, the main source of economic and political power lies in Peninsular Malaysia due to historical reasons, the level of development, and concentration of population (Khoo 2013). Second, the political parties operating in the Borneo states are to a certain extent autonomous from those in the peninsula (Noor 2013). Thus, there is a distinct possibility that whichever party or coalition wins in the peninsula can persuade politicians who have won on opposition party tickets in Sabah and Sarawak to join their coalition (Yusoff 2001). Finally, the ethnic composition of Sabah and Sarawak is quite different from that of Peninsular Malaysia. In particular, the ethnic makeup of Sabah and Sarawak is much more diverse (Sim 2010). Therefore, ethnic factors exercise a much stronger influence on the electoral outcomes in the peninsula relative to Sabah and Sarawak. Of the 222 federal constituencies, 165 (or nearly 75 percent) are from Peninsular Malaysia. It is for these reasons that this study is confined to the peninsula for the remainder of this article.

Malaysia is ethnically a very diverse nation. The indigenous ethnic groups are classified as Bumiputera (which can be transliterated as "son of the soil") and this classification includes the Malays, Orang Asli (who are the aboriginal people in Peninsular Malaysia), and the various indigenous ethnic groups in the Borneo states of Sabah and Sarawak. According to the 2010 census, the Bumiputera community made up 67.4 percent of the total Malaysian population. The other two major ethnic groups are the Chinese and Indians, who made up 24.6 and 7.3 percent of the population, respectively.

Table 1 presents the population distribution by ethnic groups from 1957 to 2010 in Peninsular Malaysia. As can be seen from the table, at the time of independence in 1957, there was almost equal representation of Malays[4] and non-Malays. Therefore, there was considerable fear among the Malays that they would be swamped by the Chinese and In-

Table 1 Population Distribution by Ethnic Group in Peninsular Malaysia, 1957–2010

	1957	1970	1980	1991	2000	2010
Malays/Bumiputera	49.8	52.7	55.3	58.3	62.4	64.6
Chinese	37.2	35.8	33.8	29.4	27.4	25.9
Indians	11.1	10.7	10.2	9.5	9.5	8.9
Others	2.0	0.8	0.8	2.7	0.7	.06
Total[a]	100.1	100.0	100.1	99.9	100.0	100.0

Sources: Leete (1996) for 1957 to 1991 data; Saw (2007) for 2000 data; and Department of Statistics (2011) for 2010 data.
Note: a. Totals do not all add up to 100 due to rounding.

dian immigrants, and their descendants (Abdullah 1997).[5] To partly assuage these fears, the British colonial authorities, together with the local political elites, ensured that the special position of the Malays and other indigenous people was enshrined in the proposed constitution for the new nation by designating Malay as the national language, Islam as the official religion, and a quota for Bumiputeras in public sector jobs and scholarships for higher education, among other items.

The Electoral System and General Elections

Malaysia has gone through thirteen general elections since independence. Table 2 provides a summary of the election results for all thirteen general elections. The last column measures the distortion effect, which is defined as the difference between the percentage share of seats that BN won and the percentage of popular votes it received.[6] The Alliance/BN benefited significantly from the distortion effect during the first eleven general elections.[7]

There were several reasons for this success. First, the rapid economic growth that Malaya, and later Malaysia, enjoyed during the first forty years of independence uplifted the livelihood of most Malaysians. In particular, universal education and health care were provided at very low cost to all citizens. Physical infrastructure improved tremendously. This track record of economic prosperity and the ensuing political and ethnic stability enabled BN to claim credit for this growth (Mutalib 2000).

Another reason for BN's success was the divisions among the opposition parties. From 1957 to 1998, the opposition was made up of fragmented individual parties choosing to go to the polls without any formal electoral pact (Ufen 2009). By operating as a multiethnic coali-

Table 2 Summary of Parliamentary Election Results for Peninsular Malaysia

General Election Number	Year	No. of Seats Won			Share of Seats Won (%)		Share of Votes Won (%)		Distortion Effect (%)
		Total	Alliance /BN	Opp.	Alliance/BN	Opp.	Alliance/BN	Opp.	
1	1959	104	74	30	71.2	28.8	51.8	48.2	19.4
2	1964	104	89	15	85.6	14.4	58.5	41.5	27.1
3	1969	104	67	37	64.4	35.6	48.6	51.4	15.8
4	1974	114	104	10	91.2	8.8	61.5	38.5	29.7
5	1978	114	94	20	82.5	17.5	57.1	42.9	25.4
6	1982	114	103	11	90.4	9.6	61.3	38.7	29.1
7	1986	132	112	20	84.8	15.2	58.1	41.9	26.7
8	1990	132	99	33	75.0	25.0	55.5	44.5	19.5
9	1995	144	123	21	85.4	14.6	66.3	33.7	19.1
10	1999	144	102	42	70.8	29.2	55.4	44.6	15.4
11	2004	165	147	18	89.1	10.9	63.7	36.3	25.4
12	2008	165	85	80	51.5	48.5	49.6	50.4	1.9
13	2013	165	85	80	51.5	48.5	45.7	54.3	5.8

tion, Alliance/BN was able to occupy the middle ground in the ethnic spectrum and not portray itself to be an ethnic-based party. However, opposition politics in Peninsular Malaysia has been dominated by two ethnic-based parties: Parti Islam Se-Malaysia (PAS), an Islamist party drawing its support almost exclusively from the Malay-Muslim community, and the Democratic Action Party (DAP), which gets most of its support from the non-Malays. This fragmentation, together with the first-past-the-post system, enabled BN to win a large number of seats comfortably, even though its share of votes was between 50 and 65 percent for each of the first eleven elections. In fact, the distortion effect mentioned earlier was amplified mainly due to the ability of BN to hold the middle ground and win convincingly in ethnically mixed seats.

A third reason for BN's long string of victories was the increasing weightage given to rural and ethnically mixed constituencies over several delineation exercises (Lim 2002). In Malaysia, the rural areas of the country are mostly populated by Malays. Over the years, there has been constant gerrymandering by the Election Commission (EC) (Lim 2002), which is tasked to delineate the apportioned constituencies based on Article 46 of the Malaysian Federal Constitution. The constitution, however, has left it to the EC to interpret and apply important but vague and undefined terms such as "a measure of weightage," "rural," and "urban" when delineating electoral constituencies. The EC's application of rural weightage has been a source of controversy as it takes precedence over the fundamental principle of equal size (Lim 2002).

However, the situation changed dramatically since the twelfth general election (GE12) held in 2008. BN's share of popular votes in Peninsular Malaysia, which has been more than 50 percent since 1974, has declined significantly in the last two general elections. A plausible reason for this shift was the emergence of a multiethnic opposition coalition that provided an alternative vision for the country's future that all communities could accept (Leong 2012). In GE12, the former deputy prime minister, Anwar Ibrahim, who had been expelled from the government and jailed in 1998, was able to forge an electoral pact between his newly formed party, Parti Keadilan Rakyat (PKR), and the two leading opposition parties (i.e., DAP and PAS) in the peninsula. This informal group then went on to produce a common election manifesto and campaigned jointly in many parts of the country. The election result was stunning. BN suffered a severe setback in terms of both the number of seats as well as the share of popular votes. More damaging, BN lost control of five of the eleven state governments in the peninsula. Among them were the most industrialized and urbanized states of Selangor, Penang, and Perak. BN was also routed in the federal territory of Kuala Lumpur, the commercial capital and largest city in Malaysia. However, within a year, the PR state government in Perak was toppled when three of its state assembly members crossed over to BN.

Following this somewhat unexpected success in 2008, the electoral pact was transformed into a coalition called the Pakatan Rakyat (PR) or People's Alliance, made up of Anwar Ibrahim's PKR, DAP, and PAS. During the past five years, PR's component parties have been able to attract a number of young leaders from the different ethnic groups in the country. State financial coffers have been substantially improved through good governance in the four PR-governed states of Kelantan, Kedah, Penang, and Selangor (Auditor General of Malaysia 2011a, 2011b).[8] In the latter two states, there was a marked improvement in administrative efficiency in the state government machinery. In spite of the BN-controlled federal government's continuous attempt to undermine this coalition (Liow and Pasuni 2010), PR has been able to maintain its unity up to the GE13, held in May 2013. While BN still had the advantage of using the extensive federal government machinery to help its election campaign, as well as had control over the mainstream media, the emergence of Internet-based news portals and blogs leveled the playing field to some extent (Gomez and Chang 2013; Pepinsky 2009; Rajaratnam 2009).

The election campaign was very intense, because for the first time in Malaysia's electoral history there was no assurance that BN would be able to retain its parliamentary majority. Ultimately, the results were disappointing to Pakatan Rakyat. As mentioned above, while there was no change in the number of seats won by both sides in Peninsular Malaysia, there was a shift in the voting pattern. In particular, BN made gains in some rural areas while it lost few urban seats.

Ethnicity and Urbanization

The foundations of participatory political systems arose from rapid industrialization. Weber (1978) argued that modern political upheaval can be traced back to economic changes and shifting populations. Nevertheless, Lerner (1964) hypothesized that a modernizing society needed to achieve a certain minimalist threshold of urbanization before a participant society can be seen to emerge. Urbanization also results in the growth of modern interest groups whose demands and opposition to political elites lead to expansion of political communities. In summary, rapid urbanization results in a decentralization of political power (Lipset 1959).

Urbanism does not just involve growth of cities. It also leads to broader social transformation. Specifically, it leads to the mobilization of both the rural and urban electorate by politicians. Thompson (2013) argues that when there is a lack of an electoral process, rural identities and the cultural discrimination felt by people of rural origin may be a source of social unrest. A clear example of this is the prolonged political impasse ongoing in Thailand. However, when a proper electoral process is in place, a flourishing two-party system may develop. An early study by Cutright (1963) in the United States affirmed that increasing urbanization seems to be conducive for the development of a competitive party system. There have also been documented cases of changes in voting patterns due to urbanization in other developed countries such as Turkey (Shmuelevitz 1996) and Italy (Fried 1967).

Urbanization has also been linked to political instability. Given the large influx of migrants into urban areas, the lack of planning of amenities by urban authorities often leads to an expectation gap between the dream of city life and the harsh realities that these migrants face. This in turn causes relative deprivation and social-psychological maladjustment, which leads to political radicalization and support for protest movements where grievances are aired through demonstrations. Such scenarios are prevalent not only in Latin America (Cornelius 1969) but also in Africa as well. Resnick (2012) classifies this marginalized group as the urban poor and found that this group tends to be politically mobilized through the use of populist strategies by opposition parties in a number of African countries.

Compared to the literature on urbanization, there has been more research on the role of ethnicity in the determination of electoral outcomes. Ethnicity is a frequently used tool to galvanize political support. Its use is especially prevalent when there is a sudden democratic transition (Ottaway 1999). In such situations, voters may resort to voting along ethnic lines as they may feel likelier to receive greater benefits by voting for a politician of the same ethnicity, rather than for someone outside of the ethnic group (van de Walle 2007). This so-called cognitive shortcut is more prevalent among those with lower education attainment and as such it may be difficult for those individuals to distinguish the differences between political parties (Noris and Mattes 2003).

The question of whether ethnicity remains a significant factor in the urban context has been a subject of debate among political analysts. One school of thought maintains that urbanization actually leads to greater interethnic competition over scarce, but highly visible, resources and opportunities (Bates 1983; Melson and Wolpe 1970).

Urbanization, however, could contribute to more cosmopolitan worldviews that may nullify the ethnicity effect (Lipset 1959; Parsons 1975). This could lead to what Thompson (2013) refers to as an "urban cosmopolitan chauvinism" bias whereby urbanites perceive themselves as far superior to their rural peers. Also, prolonged interaction with other ethnic groups increases awareness of commonly shared characteristics, whereas infrequent contact can reinforce hostile ethnic stereotyping (Allport 1979).

Prior studies (Feagin 1972; Mohd Fuad et al. 2011) have analyzed the effects of ethnicity and urbanization independently and researchers have tended to treat both factors as additive rather than interactive. In contrast, Whitby (1985) introduced an interaction term consisting of both these factors to analyze the voting patterns of politicians elected to the US Congress. He concluded that the interaction between urbanization and an increasing proportion of the black population led to more liberal voting behavior on legislation passed by the US Congress. Urbanization therefore conditions the effect of race.

To understand the role of ethnicity and urbanization in determining the electoral outcome in the context of a developing country, one has to examine the demographic development of Malaysia since its independence in 1957. During British colonial rule and immediately after independence, agriculture was the dominant sector of the economy. During that period, most urban centers in the country had large Chinese majorities. Given that most Malays were employed in the agricultural sector, they lived predominantly in rural areas. Since independence, Malaysia has seen rapid modernization and industrialization with manufacturing taking over from agriculture as the main contributor to the economy. Rapid industrialization gave impetus to the movement of population from rural villages to towns and cities. According to Yaakob, Masron, and Fujimaki (2012), Malaysia has experienced rapid growth in urbanization during the last five decades, but urbanization was more pronounced during the 1980s and 1990s. Levels of urbanization have grown from 28.4 percent in 1970 to 71 percent in 2010 (Department of Statistics 2011). Malays now form the majority in most urban centers in Peninsular Malaysia because of the rural-urban migration that has taken place (McGee 2011).

Following a racial riot that occurred after the 1969 general election, the government began the implementation of the New Economic Policy (NEP), a legislated affirmative action policy meant to rectify the economic imbalance between the Malays and the non-Malays (Chin 2001). The implementation of the NEP aided by rapid industrialization and eco-

nomic development has resulted in the rise of a Malay middle class that is increasingly less dependent on affirmative action but more interested in universal issues like participatory democracy, justice, and human rights (Saravanamuttu 2001).

Middle-class Malays are now better educated, better informed, with better access to alternative media (Pepinsky 2009). They are now more economically independent and located in urban areas and are at ease with interethnic economic and social relationships (O'Shannassy 2009). They, together with their non-Malay counterparts, have been pressing the cause for democracy more fervently under a nascent multiracial platform (Liow 1999). This cohort opted for change and is located in urban areas (Welsh 2013). The rise of the middle-class Malays culminated in the significant loss of BN's two-thirds majority in the 2008 and 2013 elections, respectively, as seen in Table 2 above (Khoo 2013).

Several factors have also contributed to the rising discontent among Malaysians across racial divisions. These include rising crime, corruption scandals, weakness of the judicial system, and increasing cost of living in general (Moten 2009). Other studies using a sociological approach argue that economic growth and the pressures of materialism and urbanization have strengthened rationalism and weakened ethnic considerations in deciding whom to vote for (Mansor 1992, 1999).

Indeed, as a country moves into the middle-income range during economic transition, it leads to changes in the social structures, beliefs, and culture that foster democracy. Huntington (1991) refers to this as a "Third Wave" of democratization. The recent election results have also triggered debate of whether we have indeed seen the end of ethnic politics in Malaysia. It is without doubt that parties canvassing on a pure ethnic political platform have been severely weakened, but to say that ethnic politics have been totally extinguished in Malaysia would be premature (Arakaki 2009; Balasubramaniam 2006; Lian and Appudurai 2011; Moten 2009; Noh 2014; Pepinsky 2009).

Nevertheless, the uncoupling of ethnic interests from ethnic identity of their advocates is steadily progressing as exemplified by the actions undertaken by opposition parties in power at the state level. Politicians from both sides of the political divide are beginning to see the benefits of addressing sensitive or racially charged issues as an across-the-board responsibility (Mohamad 2008). This is evidenced in the last two elections with the emergence of a two-party coalition system with the charismatic Anwar Ibrahim leading a credible opposition coalition that has presented itself as an alternative to the ruling BN coalition (Mohamad 2008; Moten 2009). The emergence of a two-party coalition system may be linked to the rapid urbanization of the country.

Thompson (2013), however, asserted that although the current discursive argument points to a sharp urban-rural divide between city and village dwellers, it ignores the fact that there is substantial movement between rural and urban voters. He points out that although PAS used the pervasive urban cosmopolitan chauvinism tactic in mobilizing the rural dwellers against its arch-enemy BN, it has not been an outright success. In contrast to the Mahathir administration, which has been accused of neglecting rural development (Thompson 2013), the former prime minister Abdullah Badawi and the current prime minister Najib Razak have made the interests of rural Malaysia a key cornerstone in their policymaking decisions. They realize that there is much to lose if rural Malaysia continues to be ignored in the name of economic progress.

Increasing urbanization has affected the ethnic composition of urban areas in Malaysia. Thus, the question is, which of these two factors is leading to the changes in the electoral dynamics of Peninsular Malaysia? An econometric methodology to separate and quantify the marginal effect of the urbanization-ethnicity interaction is presented in the next section.

Data and Methodology

Data

The dependent variable in the model is defined as the proportion of votes won by BN in each of the 165 parliamentary constituencies considered in this study, calculated as the number of votes that BN garnered for a parliamentary constituency, divided by the total number of valid votes for that corresponding parliamentary constituency. These data were obtained from the electoral results published online by the Malaysian Election Commission.[9]

The explanatory variables in the model specification include (1) the total number of eligible voters from each of the four ethnic groups in each parliamentary constituency, classified as Bumiputera, Chinese, Indians, and Others; and (2) Area—the physical area of the parliamentary constituencies that is used as a continuous measure of urban development. The ethnic group classification of Bumiputera encompasses the Malays, Orang Asli (which refers to the aborigines of Peninsular Malaysia), Bumiputera Sabah, and Bumiputera Sarawak.

Using *The Star* newspaper[10] (April 21, 2013), which had obtained data from the Malaysian Election Commission on the percentage composition of the four ethnic groups in each of the 165 parliamentary constituencies, the total number of eligible voters from each of the four ethnic groups was computed by multiplying the respective ethnic percentage composition with the total number of eligible voters in that particular constituency. This study does not directly use the percentage composition of the four ethnic groups provided by *The Star* as the explanatory variables primarily because the use of such compositional explanatory variables (with a unit sum constraint) in a regression analysis, even for a linear regression model, may lead to biased and incorrect estimates, resulting in misleading inferences (Hron, Filzmoser, and Thompson 2012).[11]

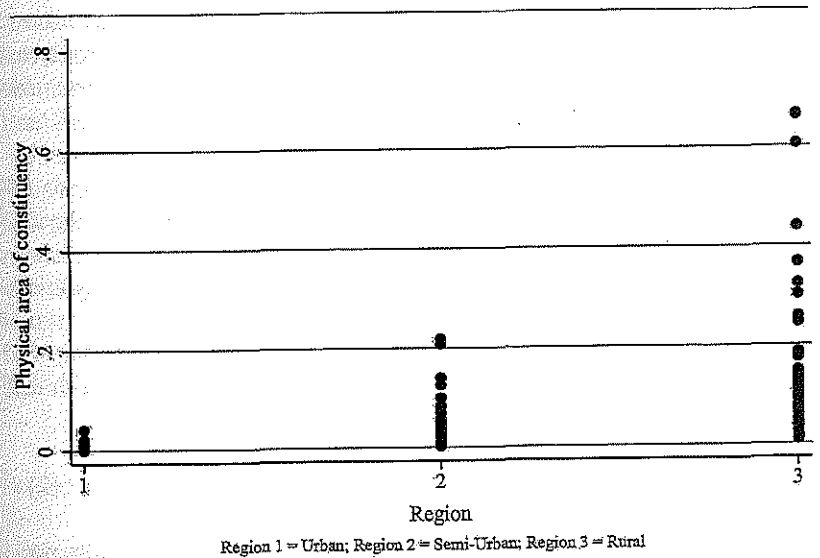
Data on the physical area of the parliamentary constituencies were obtained from Greenberg and Pepinsky (2013), and have also been used by Ostwald (2013) in his analysis of the relationship between district size and voter density across the Malaysian parliamentary constituencies.

In addition to using Area as the continuous measure of urbanization, we also use data from Politweet that categorize the 165 parliamentary constituencies as either urban, semi-urban, or rural to aid our analysis.[12] According to Politweet, a constituency is defined as rural if there are villages, small towns, or farmlands distributed within the constituency. Semi-urban constituencies are those that contain larger towns and/or numerous small towns as compared to a rural constituency, and may also contain small villages. Urban constituencies refer to cities where urban development occurs in a majority of the constituency. The Politweet classification was done based on Google Maps satellite imagery and the Election Commission of Malaysia maps. The three-step methodology in determining the level of urban development of a constituency can be found in Appendix 1. Based on this classification, the number of rural, semi-urban, and urban constituencies is eighty-one, forty-four, and forty, respectively.

By combining the data on the physical area of the constituencies with the Politweet urbanization classification, Figure 1 plots the areas of the 165 parliamentary constituencies in their respective urban development classification, with the areas ranging from 0.0005 to 0.66433.[13]

It can be seen from Figure 1 that urban constituencies are associated with smaller areas, while rural constituencies are associated with larger areas. In particular, the median areas for the rural, semi-urban, and urban constituencies are 0.060, 0.037, and 0.003, respectively. Therefore, it is

Figure 1 Areas of the 165 Parliamentary Constituencies in Their Respective Urban Development Classification



Region 1 = Urban; Region 2 = Semi-Urban; Region 3 = Rural

reasonable to conclude that parliamentary constituencies that are larger in physical size are also more rural in nature, while parliamentary constituencies that are smaller in physical size are more urban in nature. In other words, along the area continuum, constituencies with small areas refer to urban constituencies while constituencies with large areas refer to rural constituencies.

Methodology

The dependent variable, proportion of votes garnered by BN, is a proportion quantity restricted to an interval of 0 and 1 (or referred to as a unit interval). Therefore, the model specification for a fractional dependent variable, y_i, as described by Papke and Wooldridge (1996), is assumed as follows:

$$y_i = F(x_i' \beta) + u_i, i = 1,2,\ldots,n \quad (1a)$$

$$E(y_i|x_i) = F(x_i' \beta), i = 1,2,\ldots,n \quad (1b)$$

where x_i are the explanatory variables for the ith observation and $F(.)$ is typically chosen to be a known cumulative distribution function satisfying $0 < F(x_i'\beta) < 1$ for all $x_i'\beta \in R$, ensuring that the predicted values of y will lie between the unit interval. The term u_i is a random error term with a conditional mean of 0. Equation 1a decomposes the observation into two components—the deterministic (or predictable) component given by $F(x_i'\beta)$ and a random (or unpredictable) component given by u_i. Equation 1b is derived by taking the conditional expectations of Equation 1a, and it models the conditional expectation of y_i (i.e., the average value of y_i given a set of x_i values) as a deterministic function given by $F(x_i'\beta)$. It is the conditional expectation function in Equation 1b that is estimated.

The vector of parameters in Equation 1a, β, is estimated using the Bernoulli quasi-maximum log-likelihood estimator (QMLLE), with the Bernoulli log-likelihood function given by

$$l_i(\beta) = y_i log[F(x_i'\beta)] + (1 - y_i) log[1 - F(x_i'\beta)] \quad (2)$$

The QMLLE of β in this model is consistent and asymptotically normally distributed irrespective of the conditional distribution of y_i. In this article, $F(.)$ is chosen to be a logistic function that is strictly monotonic and defined as

$$F(z) \equiv \frac{\exp(z)}{[1 + \exp(z)]}, \quad (3)$$

where z is defined as $x_i'\beta$ from Equation 1a. The model specification in Equation 1a, coupled with this logistic function in Equation 3, is hence referred to as the fractional response logit model.[14]

The nonlinear model specification in Equation 1a is chosen over the linear specification because linear regression models are not appropriate for bounded values. Therefore, using the linear regression model will result in model misspecification with inefficient estimators, thus leading to incorrect inferences (Hron, Filzmoser, and Thompson 2012). Moreover, predicted values from a linear specification are not guaranteed to lie in the unit interval. Linear regression techniques also assume that the effect of any explanatory variable is constant throughout the range of values of the explanatory variable.

Further, by specifying the fractional response model in Equation 1a, no a priori assumption on the range of values that the dependent variable can take is made, except that it must be within the unit interval.[15] In doing so, the model specified in Equation 1a accounts for all possibilities, including the possibility of a constituency observing a value of either 0 or 1 for the dependent variable. It should also be noted that the fractional response model is used not only when the dependent variable records extreme values of 0 or 1, but that it has an added benefit of being able to deal with such values, avoiding the previous practice of ad hoc transformations.

Model Specification

Using Equation 1b, the expected proportion of votes garnered by BN is modeled as a function of the number of voters in each of the four different ethnic groups, and the area of the parliamentary constituencies that is used as a continuous measure of urban development. This model is further augmented by including interaction terms between the variables. The model is subsequently represented as such:

$$E(y|x) = F\begin{pmatrix} \beta_0 + \beta_1 Bumiputera + \beta_2 Chinese + \beta_3 Indians + \beta_4 Others + \\ \beta_5 Area + \beta_6 Bumiputera.Area + \\ \beta_7 Chinese.Area + \beta_8 Indians.Area + \beta_9 Others.Area \end{pmatrix} \quad (4)$$

where Bumiputera, Chinese, Indians, and Others represent the number of voters (in 10,000s) from the respective ethnic groups of the parliamentary constituencies, and Area is the proxy to measure the level of urban development for the parliamentary constituencies.

The present study improves earlier studies (Feagin 1972; Mohd Fuad et al. 2011) by including the interaction effects of urbanization on ethnicity in determining the proportion of votes garnered by BN. The inclusion of interaction terms accounts for the synergy effect of both ethnicity and urbanization, thereby allowing for the complementary effect of urbanization on ethnicity to influence the proportion of votes garnered by BN. Moreover, Brambor, Clark, and Golder (2006, 64) note that when analyzing political studies, interaction terms should be included whenever the analysis involves a conditional hypothesis, defined as when "a relationship between two or more variables depends on the value of one or more other variables."

Results

Estimation Results

The model specified by Equation 4 is estimated by quasi-maximum likelihood methods and the results are presented in Table 3.

Table 3 Fractional Response Logit Regression Results

Variable	Regression Results
Bumiputera	−0.0014
	(0.0237)
Chinese	−0.2279***
	(0.0284
Indians	0.0034
	(0.0727)
Others	−1.2622***
	(0.4859)
Area	−0.0789
	(1.0256)
Bumiputera x Area	0.0107
	(0.2349)
Chinese x Area	1.5269**
	(0.6691)
Indians x Area	−1.3102
	(1.3504)
Others x Area	6.0453
	(3.9607)
Constant	0.3518***
	(0.1163)

Notes: As discussed in the text, only the sign and not the magnitude of the coefficients presented in Table 3 can be interpreted. Standard errors are in parentheses. *$p \leq 0.10$; **$p \leq 0.05$; ***$p \leq 0.01$

Before interpreting the results, a few points are in order with regard to the coefficients. First, for a fractional response logit model, unlike linear regression models, β_j alone does not represent the marginal effect of the explanatory variable x_j. Appendix 2 details how the marginal effects of an explanatory variable of interest can be computed. Second, the sign of the coefficient β_j for this class of models indicates the direction of the marginal effects of the explanatory variable on the dependent variable. Taking the above two points together, only the signs of the coefficients in Table 3, and not the magnitudes, are interpretable.

For the purpose of this article, the discussion of results will be centered on the three main ethnic groups that are the most important in Peninsular Malaysia politics: the Bumiputera, Chinese, and Indians. Referring to Table 3, the results show that of the three main ethnic groups, only the *Chinese* ethnic group is a statistically significant variable, bearing a negative coefficient. *Area* was found to be statistically insignificant at the conventional level of significance. This result implies that the degree of urban development of a particular constituency, by itself, does not influence the level of support for the government. However, the positive and statistical significance of the interaction term between *Area* and *Chinese* reveals that there is a complementary effect between these two variables. The results imply that there is a significant extra effect of urbanization (as proxied by *Area*) on the Chinese voters to influence the proportion of votes garnered by BN, ceteris paribus. More specifically, the positive coefficient of this interaction term implies that for every additional Chinese voter in a parliamentary constituency, ceteris paribus, the increase in the proportion of votes to BN is higher in larger (rural) parliamentary constituencies than smaller (urban) parliamentary constituencies. This positive effect negates the decrease in the proportion of votes to BN arising from the negative coefficient of *Chinese*.

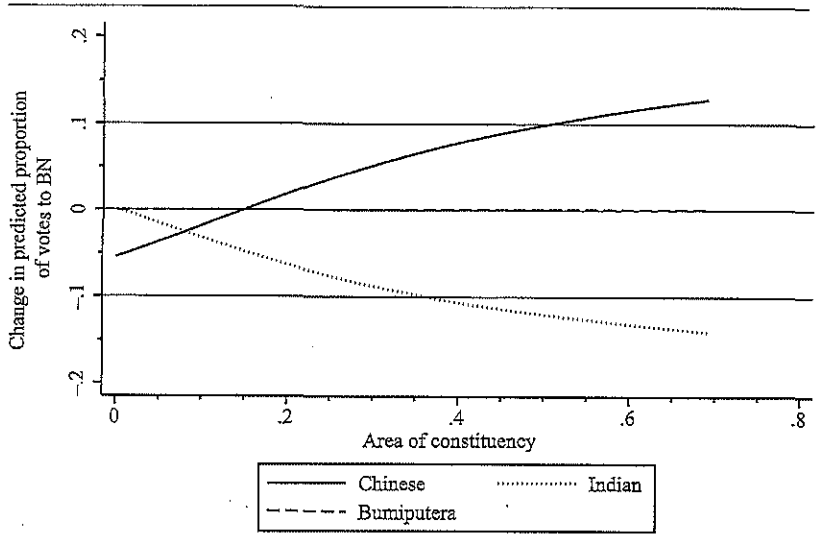
However, the other ethnic group variables (*Bumiputera* and *Indians*), and their interaction with *Area*, are all statistically insignificant. This implies that an increase in either Bumiputera or Indian voters in a parliamentary constituency, ceteris paribus, will result in no change in the proportion of votes to BN. There is also no added effect that comes from the level of urbanization in that constituency. Therefore, it seems that the voting intentions of the Bumiputeras and Indians are constant irrespective of the level of urban development of the parliamentary constituency that they are in.

The results suggest that the *Chinese-Urbanization* factor is having the most dominant influence on the proportion of votes garnered by BN. The next subsection provides plots to visualize the underlying patterns in the ethnicity and urbanization effects on the proportion of votes to BN. These plots provide interesting insights into understanding the underlying relation between these variables and reveal deviations from patterns that may not be detected easily.

Ethnicity and Urbanization Effects

Using the estimation results in Table 3, Figure 2 provides a visualization of the ethnicity and urbanization effects by plotting the *marginal effects* of the number of voters from each ethnic group on the predicted proportion of votes to BN, across the range of areas of the parliamentary constituencies. Specifically, Figure 2 plots the *change* in the predicted proportion of votes to BN that arises from an *increase* in the number of voters from each respective ethnic group, for the range of areas of the parliamentary constituencies. The horizontal axis in Figure 2 represents

Figure 2 Marginal Effects of Respective Ethnic Voter Numbers on the Predicted Proportion of Votes to BN, Across the Range of Constituency Areas



the range of areas of the 165 parliamentary constituencies and is used as a continuous measure of urban-rural development, where a constituency with a smaller area is associated with being an urban constituency and vice-versa. The vertical axis represents the change in the predicted proportion of votes to BN that arises from an increase in the number of voters from each of the ethnic groups.

Referring to Figure 2, it is evident that regardless of the area of the constituency, an increase in the number of Bumiputera voters in a constituency, ceteris paribus, results in no change in the predicted proportion of votes to BN. Therefore, whether the constituency is an urban or rural region, an increase in the number of Bumiputera voters in that constituency, ceteris paribus, does not alter the level of support for the ruling coalition, BN.

Furthermore, the marginal effects of the Chinese voters are significantly larger than those of the Bumiputera voters. Given an increase in the number of Chinese voters relative to the other ethnic groups, the change in the predicted proportion of votes to BN is negative in smaller (urban) constituencies while this change is positive in larger (rural) constituencies. Although the support for BN is positive in large (rural) constituencies with a higher relative number of Chinese voters, this level of rural Chinese support dwindles with larger constituencies, as shown by the tapering of the solid line in Figure 2. This implies that the level of urban development of a constituency influences the level of Chinese support for BN.

The predicted change in proportion of votes to BN from an increase in the number of Indian voters, ceteris paribus, is negative for all areas of constituencies in Figure 2. However, these marginal effects are statistically insignificant from zero and therefore an increase in the relative number of Indian voters in either an urban or rural constituency will not influence the level of support for BN.

Ethnicity and Urbanization Effects: An Extended Analysis Using the Categorical Measure of Urban Development

The earlier analysis used area of constituencies as the continuous measure of urban development and subsequently plotted marginal effect graphs to illustrate the ethnicity and urbanization effects on the proportion of votes to BN. To extend and complement the former analysis, we also show the ethnicity and urbanization effects on the proportion of votes to BN by using the *categorical* measure of urban development where, as discussed earlier, the parliamentary constituencies are classified as urban, semi-urban, or rural. Specifically, for each ethnic group, we plot and observe the predicted proportion of votes to BN in urban, semi-urban, and rural constituencies, over the range of number of voters from that ethnic group. To do this, we substitute, in turn, the median areas representing each of the region classifications (0.003 for urban, 0.037 for semi-urban, and 0.060 for rural) into Equation 1b and obtain the predicted proportion of votes to BN for the range of voter numbers in each ethnic group. Figures 3–5 subsequently display the results.

In Figure 3, the predicted vote share for BN across the three region classifications declines very slightly throughout the range of Bumiputera voter numbers. Therefore, since an increase in the number of Bumiputera voters, ceteris paribus, changes the predicted proportion of votes to BN only marginally in either an urban, semi-urban, or rural constituency, it can be inferred that for the Bumiputeras, ethnicity has little influence on the level of support for BN. It is notable that in all three constituency types, the Bumiputera predicted vote to BN is less than 50 percent. However, it is evident from Figure 3 that it is the level of urban development that influences the level of support for BN among the Bu-

Figure 3 Predicted Vote Share for BN in Urban, Semi-Urban, and Rural Constituencies for Varying Numbers of Bumiputera Voters

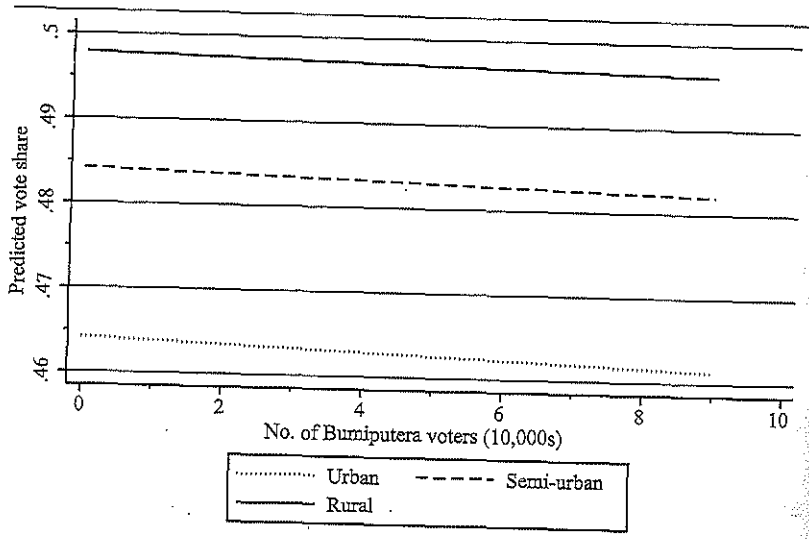
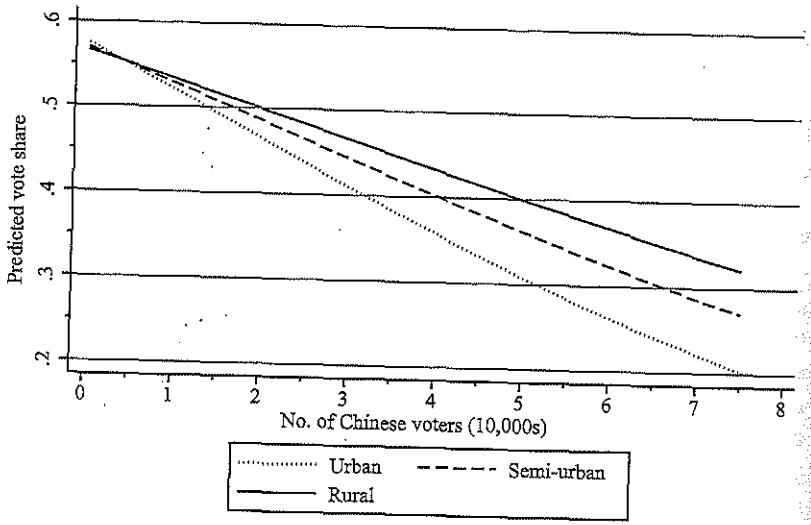


Figure 4 Predicted Vote Share for BN in Urban, Semi-Urban, and Rural Constituencies for Varying Numbers of Chinese Voters



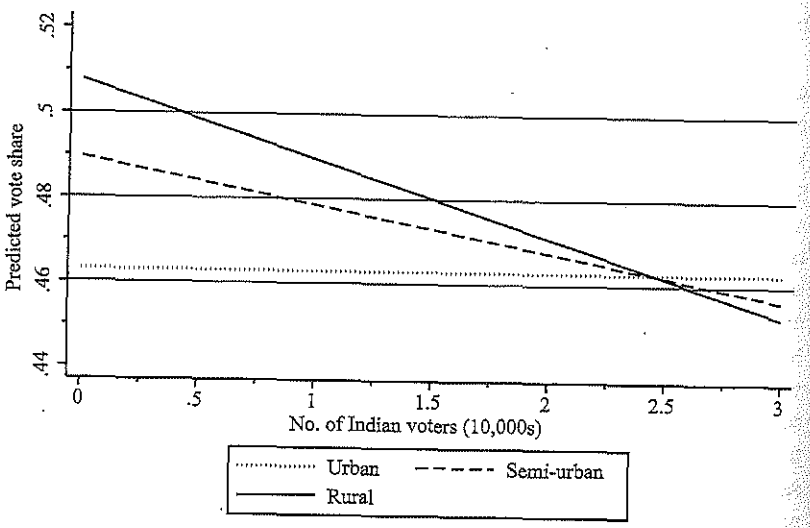
miputeras. Regardless of the number of Bumiputera voters in a given constituency, the model predicts that the proportion of votes to BN will be the lowest in an urban constituency, and it will be the highest in a rural constituency.

Figure 4 plots the predicted vote share to BN across the three region classifications over the range of Chinese voter numbers, revealing the presence of both ethnicity and urbanization effects. At small Chinese voter numbers (around 5,000) in a given constituency, the predicted proportion of Chinese votes to BN is approximately 55 percent, regardless of whether the constituency is urban, semi-urban, or rural. However, the negative trend lines for all three region classifications show that the level of support for BN decreases quite rapidly as the number of Chinese voters increases, thus capturing the ethnicity effect. This ethnicity effect is further fueled by the urbanization factor, as seen by the divergence in the predicted proportion of votes to BN across the three region classifications, with urban constituencies reflecting a sharper drop in the proportion of votes as compared to rural constituencies. In particular, for the same increase in the number of Chinese voters from 5,000 to 75,000, the predicted proportion of votes to BN in an urban constituency drops from 55 percent to approximately 20 percent, as compared to the drop to approximately 33 percent for a rural constituency.

In contrast to the divergent trend lines in Figure 4 for the Chinese voters, Figure 5 shows the trend lines across the three region classifications for the Indian voters converging. In a constituency with a low number of Indian voters, the predicted proportion of votes to BN is the highest if it is a rural district, and is the lowest if it is an urban district. This disparity in the vote share to BN across the three region classifications diminishes in constituencies with a higher number of Indian voters, with the predicted vote share to BN decreasing for both the rural and semi-urban constituencies. The predicted vote share to BN for an urban constituency, however, remains constant over the range of Indian voters. It is also noted that for constituencies with fewer than approximately 25,000 Indian voters, the predicted proportion of votes to BN is the lowest in urban areas, reflecting the urbanization effect on the support for BN.

However, the findings from Figure 5 for Indian voters should be treated with some caution. In most constituencies, the Indian share of votes is relatively small. The Kota Raja parliamentary constituency has the highest number of Indian voters with a 29.5 percent share. Another nine parliamentary constituencies have 20–24 percent of Indian voters, while in fifteen other parliamentary constituencies, they form 10–20 per-

Figure 5 Predicted Vote Share for BN in Urban, Semi-Urban, and Rural Constituencies for Varying Numbers of Indian Voters



cent of the voters. Thus out of 165 parliamentary seats in Peninsular Malaysia, Indian vote share exceeds 10 percent only in twenty-five parliamentary seats, thus reducing their impact on the electoral outcome in most parliamentary seats.

A few additional points are also observed from Figures 3, 4, and 5. First, across all the ethnic groups, the support for BN is always the lowest in urban constituencies and the highest in rural constituencies. This is a clear indication that the level of urban development influences the level of support for BN, with more urbanized constituencies being pro-opposition. Second, as shown in Figure 3, the Bumiputera vote is evenly split with a slim majority for the opposition Pakatan Rakyat. In the aggregate, Bumiputera support for BN ranges from 46 percent in urban areas to 49 percent in rural areas. Third, the ethnic effect is the strongest among the Chinese, with support for BN declining sharply in constituencies that have a larger number of Chinese voters relative to the other ethnic groups.

Discussion

GE13 was arguably the most closely contested election in Malaysian history. In terms of popular votes, GE13 was the worst outcome for the ruling coalition in BN's history. For Malaysia as a whole, PR obtained more votes than BN. In Peninsular Malaysia, BN obtained only 45.7 percent of the votes cast, while PR obtained more than 54 percent of the votes. It is also intriguing to note from Figure 3 that even among Bumiputera voters, BN was unable to obtain majority support. Among the non-Bumiputera, a clear majority voted for PR. Yet, BN won eighty-five seats to PR's eighty seats. A large proportion of BN's electoral success, however, can be attributed to victories in rural parliamentary seats as classified by Politweet. Of the eighty-five seats BN won in Peninsular Malaysia, sixty-six were classified as rural.

The key to understanding this outcome is to examine Chinese voter behavior in constituencies where their numbers are small. As can be seen from Figure 4, when the Chinese voters make up a small number in absolute terms in urban, semi-urban, and rural constituencies, their support for BN is more than 50 percent. The higher than 50 percent support for BN by Chinese voters appears to offset the slightly less than 50 percent Malay support for BN in rural areas, thus helping BN over the finish line in rural parliamentary constituencies. So, ironically, given that most of BN's eighty-five electoral seats are from the rural constituencies, BN's victory in the thirteenth general election can be partly attributed to the support of Chinese voters, and to a lesser extent, Indian voters, in constituencies where they are small in number. BN won sixty-one of the sixty-six rural seats where Chinese voters numbered 20,000 or fewer. This rural effect is further amplified by the malapportionment of voters between rural and urban constituencies. The variance between rural and urban seats has been increasing over several rounds of electoral delineations (Lim 2002; Ostwald 2013).

It is not clear why the Chinese support for BN increases when their numbers decrease in electoral constituencies (see Figure 4). One possible explanation is that where Chinese voters make up a small percentage of the electorate, they may tend to feel less secure. As their numbers increase, they may lose their sense of insecurity, and thus their support for BN declines sharply to as much as only 20 percent, especially for urban seats. Second, in rural constituencies, the role of government development spending has a bigger impact on the economic well-being of voters. This may influence their preference for BN among Chinese voters in rural areas. The motivation for the change in Chinese voter behavior as their numbers change needs to be explored in greater depth.

What are the implications of these findings for BN and PR? As the country transitions from a developing nation to a developed nation, Malaysia will become increasingly urbanized. Rural and semi-urban areas will become increasingly more urban in character as development

moves beyond urban areas. This may result in BN's vote share decreasing across all ethnic groups. It is also compounded by the fact that younger voters tend to read news on the alternate media over which the BN government has very little control. Future delineations that skew rural-urban constituency sizes are becoming more difficult due to a more aware electorate.[16] Thus, BN has to adopt a strategy that accounts for these factors.

BN has to refocus on urban areas for success in future elections. BN needs to reach out to urbanites of all ethnic groups and not treat urban seats as a lost cause. Less practice of divisive politics would be one step in the right direction. They should also woo both the Malay and the non-Malay votes in rural and semi-urban areas where they have some advantage in these areas. This should be especially targeted toward the Chinese voters since their voting patterns in rural and semi-urban areas play a decisive role.

Pakatan Rakyat has a clear decisive advantage in urban parliamentary seats. Support from all ethnic groups is above 50 percent. Urban parliamentary seats are therefore the easiest for PR to retain with popular support. Therefore, the strategy PR should employ to maintain urban support would be the provision of good governance, especially efficient municipal services in states where PR is the ruling coalition (Penang, Selangor, and Kelantan). This will ensure that there will be no swing back to BN in the coming general election. Contrary to BN's less than decisive advantage in rural and semi-urban seats, PR has opportunities to make inroads in these two seat classifications. Early identification of candidates for each rural and semi-urban constituency is imperative and it would be especially advantageous if the identified candidates were local politicians. Early identification of candidates would reduce political bickering among would-be aspirants and allow identified candidates to work in their respective constituencies in building up their support base.

Conclusion

More recently, the debate of ethnicity's having an effect on voter sentiment has attracted considerable interest in a multiethnic society such as Malaysia. However, the findings in this article suggest that there is more to it than ethnic undertones. This article has made two contributions. First, although the results of Malaysia's GE13 displayed an ethnic effect, the rapid urbanization of the country has also played a role in determining the outcome of the elections. Malaysians across ethnic lines voted overwhelmingly for PR in urban areas, whereas support for BN remains

reasonably strong in rural areas. This article also sheds new light on the long-standing explanation that BN's continued dominance in Malaysian general elections is due to the solid rural Bumiputera vote. Our analyses have shown that non-Bumiputera votes in rural and semi-urban areas were the key to BN's holding on to power despite having slightly less than 50 percent support from Bumiputera voters, even in rural constituencies. However, reliance on the non-Bumiputera voter group is highly precarious due to the shrinking number of Chinese voters in rural areas. This is caused by rural-urban migration and the decline in birthrates among the Chinese population (Department of Statistics 2012).

Second, the fact that Malaysia is becoming increasingly urbanized also needs to be taken into account. Political parties from both sides of the political divide need to take heed of this development. As the country develops further toward becoming a high income nation, voting patterns for race-based parties may decline significantly. Instead, political parties that can identify with the rising demands of voters on bread and butter issues like the increasing cost of living, social justice, corruption, and basic human rights may gain the upper hand in future general elections.

A cautionary remark, however, is in order where the main point of contention would lie in the categorization of seats into urban, semi-urban, and rural constituencies done by Politweet. They have acknowledged that a better method can be used for the categorization of parliamentary constituencies through a gridding process (Balk et al. 2006). This study looked at the impact of urbanization and ethnicity on electoral support at the parliamentary level in aggregate. Local micro-factors that may have influences on the results of individual seats were not considered in our analysis. Therefore, future studies may look at the impact of urbanization and ethnicity at the state legislative level, which may provide further insight to the results obtained at the federal level.

Jason Wei Jian Ng is lecturer with the Department of Econometrics and Business Statistics at Monash University Malaysia and can be contacted at jason.ng.wj@monash.edu. He received his PhD in econometrics from Monash University Australia. He is an early career researcher whose research focuses on applied econometrics in a wide range of policy issues such as financial development and economic growth, subjective well-being, and residential house pricing.

Gary John Rangel is lecturer of finance at Monash University Malaysia and can be contacted at gary.rangel@monash.edu. He received his PhD in finance from University Science Malaysia. His research focuses on behavioral finance, real estate, risk perception, and Malaysian politics and his work has appeared in *International Finance Review* and *International Journal of Banking and Finance*.

Santha Vaithilingam is associate professor at Monash University Malaysia and is head of the Econometrics and Business Statistics Department. She can be contacted at santha.vaithilingam@monash.edu. Her research interests are in computable general equilibrium (CGE), financial econometrics, and behavioral economics. More recently, her work has been focused on modeling human behavior pertaining to technology adoption using advanced econometric techniques.

Subramaniam S. Pillay is associate professor of business and finance at Taylor's University, Malaysia, and can be contacted at subramaniam.pillay@taylors.edu.my. He received his PhD in policy analysis from the University of British Columbia. His research focuses on economic and currency crisis, and politics in Southeast Asia.

Appendix 1

The following three-step methodology was employed by Politweet in identifying the level of urban development of the parliamentary constituencies.

1. Locate the constituency on Google Maps (and Bing Maps, when the image was not clear).
2. Identify the area covered by urban development, and the degree of development.
3. Define the seat as either rural, semi-urban, or urban based on the definitions of the classifications described in the text.

Appendix 2

Using the general model specification in Equation 1, the marginal effect of an explanatory variable of interest, x_j, on the conditional mean $E(y|x)$ is given as

$$\frac{\partial E(y|\boldsymbol{x})}{\partial x_j} = f(\boldsymbol{x}_i'\boldsymbol{\beta}).\beta_j \tag{A2.1}$$

where $f(x_i'\beta)$ is the derivative of the cumulative logistic distribution function defined as

$$f(z) \equiv \frac{dF(z)}{dz} = \frac{\exp(z)}{[1+\exp(z)]^2} \tag{A2.2}$$

and β_j is the coefficient attached to the jth explanatory variable. Therefore, to compute the value of the marginal effect of x_j on the conditional expected value of y using Equation A2.1, appropriate values of all the explanatory variables, such as the mean or median value of the explanatory variables, need to be substituted into the function $f(x_i'\beta)$ before multiplying it with β_j.

Notes

The authors would like to thank Kai Ostwald for facilitating the acquisition of the data on the areas of the electoral districts; Tom Pepinsky for providing the details behind the calculation of these areas; Peter Filzmoser for his assistance pertaining to our queries on the isometric log-ratio transformation; Mahendhiran Nair for his comments on this article; and the editor and the two referees for their constructive feedback in the reviewing process of this article.

1. From 1957 until 1973, the coalition was known as the Alliance Party and was made up of three parties—the United Malays National Organisation (UMNO), the Malaysian Chinese Association (MCA), and Malaysian Indian Congress (MIC)—representing the three major ethnic groups in Peninsular Malaysia (or Malaya, as it was then known). Its name was changed to National Front or Barisan Nasional in 1971 after it expanded to become a thirteen-member coalition encompassing political parties from the Borneo states of Sabah and Sarawak, as well as those from Peninsular Malaysia.
2. The terms *constituency* and *seat* are used interchangeably.
3. In addition to the thirteen states, there are three federal territories that are directly administered by the central government: Kuala Lumpur, the main commercial center of the country; Putrajaya, the administrative capital; and Labuan Island off the coast of Sabah.
4. In Peninsular Malaysia, Malays form nearly 98 percent of the Bumiputera population; so for all intents and purposes, Bumiputera and Malay can be used interchangeably.
5. These fears did not abate with the granting of independence by the British. Singapore's founding prime minister Mr. Lee Kuan Yew in his official autobiography (Lee 1998) indicated this fear still persisted during the formation of Malaysia in 1963. There was fear that the inclusion of Singapore in Malaysia would "swamp the Malays" (Lee 1998, 363).
6. If we include Sabah and Sarawak, BN's advantage is even bigger as the rural overweightage is more pronounced in those two states.
7. The distortion effect calculation is based on Brown (2005).
8. Penang recorded a budget surplus of RM138 million in 2011, an increase of 312 percent compared to RM33 million in 2010 when there was a 95 percent reduction in debt. Selangor also recorded an increase in revenue from RM1.57 billion in 2010 to RM1.634 billion in 2011 (Auditor General of Malaysia 2011a, 2011b).
9. The results of the thirteenth general elections for each partiamentary seat can be obtained from the Malaysian Election Commission website (http://result pru13.spr.gov.my/module/keputusan/paparan/paparan_Laporan.php#).
10. *The Star* is a daily newspaper owned by MCA, a component party of BN.
11. In particular, compositional data do not follow the usual Euclidean geometry where most statistical methods rely on it. Therefore, Hron, Filzmoser, and Thompson (2012) recommend transforming compositional variables to the Euclidean geometry via an isometric log-ratio transformation. The interested reader can refer to that paper for further details. In this study, we do

not attempt to use the isometric transformed variables because (1) there are interpretability issues (even in a linear regression setting) and (2) there is currently no research on how to estimate regressions with interaction variables that contain both compositional and noncompositional variables, as will be shown later in our model specification.

12. Politweet is a nonpartisan research firm that has been monitoring Malaysian politics and activism on Twitter since 2009, and on Facebook since late 2012. See http://politweet.wordpress.com/2013/05/21/the-rural-urban-divide-in-malaysias-general-election/.

13. The area figures do not contain any units of measurement. According to Pepinsky, the physical areas of the electoral districts were created from the Geographic Information Systems (GIS) maps and were subsequently imported into R whereby the "maptools" library was used to calculate the area of each polygon representing an electoral district. The procedure in R does not return any unit of measurement. Nevertheless, a larger number represents a larger physical area of a constituency.

14. Using either the probit or logit model is an alternative strategy, but it involves transforming the dependent variable into a binary form, and will not provide meaningful results in quantifying the ethnicity and urbanization effects on BN's electoral performance.

15. Based on the data collected, the range of the dependent variable (proportion of votes garnered by BN) is between 0.13 and 0.84. Therefore, one could argue that given that the dependent variable does not lie near the unit interval bounds, a linear regression could suffice. However, doing so will be making an a priori assumption that the range of values the dependent variable can take is a subset of the unit interval.

16. An increasingly aware electorate has culminated into nongovernmental organizations (NGOs) seeking to pressure the Election Commission to ensure a clean and fair electoral process in Malaysia. The most well-known NGO is the Coalition for Clean and Fair Elections, popularly known as BERSIH. It has held mass rallies in 2007, 2011, and 2013, demanding reform of the postal balloting system, free and fair access to mass media for all parties, cleaning up of the electoral roll, elimination of dirty money politics, as well as elimination of corruption.

References

Abdullah, Firdaus Hj. 1997. "Affirmative Action Policy in Malaysia: To Restructure Society, to Eradicate Poverty." *Ethnic Studies Report* 15, 2: 189–221.

Allport, Gordon W. 1979. *The Nature of Prejudice*. 25th anniversary ed. Reading, MA: Addison-Wesley Publishing Company.

Arakaki, Robert K. 2009. "2008 Malaysian Election: The End of Malaysia's Ethnic Nationalism?" *Asian Politics and Policy* 1, 1: 79–96.

Auditor General of Malaysia. 2011a. "Auditor General's Report 2011 on the Penang State Government Financial Statement and the Financial Management of the State Department and Agencies." Kuala Lumpur, Malaysia: National Audit Department.

———. 2011b. "Auditor General's Report 2011 on the Selangor State Government Financial Statement and the Financial Management of the State Department and Agencies." Kuala Lumpur, Malaysia: National Audit Department.

Balasubramaniam, Vejai. 2006. "Embedding Ethnic Politics in Malaysia: Economic Growth, Its Ramifications and Political Popularity." *Asian Journal of Political Science* 14, 1: 23–39.

Balk, D. L., U. Deichmann, G. Yetman, F. Pozzi, S. I. Hay, and A. Nelson. 2006. "Determining Global Population Distribution: Methods, Applications and Data." In *Advances in Parasitology*, ed. Alastair Graham, Simon I. Hay, and David J. Rogers, 119–156. Amsterdam, Holland: Academic Press.

Bates, Robert H. 1983. "Modernization, Ethnic Competition and the Rationality of Politics in Contemporary Africa." In *State Versus Ethnic Claims: African Policy Dilemmas*, ed. Donald Rothchild and Victor A. Olorunsola, 152–171. Boulder, CO: Westview.

Boo, Su Lyn. 2013. "GE13 an Urban, Not Chinese Swing, Says Analysts." *The Malaysian Insider*. www.themalaysianinsider.com/malaysia/article/ge13-an-urban-not-chinese-swing-say-analysts (accessed October 10, 2013).

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14, 1: 63–82.

Brown, Graham K. 2005. "Playing the (Non)Ethnic Card: The Electoral System and Ethnic Voting Patterns in Malaysia." *Ethnopolitics* 4, 4: 429–445.

Chin, James. 2001. "Malaysian Chinese Politics in the 21st Century: Fear, Service and Marginalisation." *Asian Journal of Political Science* 9, 2: 78–94.

Cornelius, Wayne A., Jr. 1969. "Urbanization as an Agent in Latin American Political Instability: The Case of Mexico." *The American Political Science Review* 63, 3: 833–857.

Cutright, Phillips. 1963. "Urbanization and Competitive Party Politics." *The Journal of Politics* 25, 3: 552–564.

Department of Statistics (DOS). 2011. "Population and Housing Census of Malaysia." Kuala Lumpur, Malaysia: Department of Statistics.

———. 2012. "Population Projections, Malaysia 2010–2040." Kuala Lumpur, Malaysia: Department of Statistics.

Feagin, Joe R. 1972. "Civil Rights Voting by Southern Congressmen." *The Journal of Politics* 34, 2: 484–499.

Fried, Robert C. 1967. "Urbanization and Italian Politics." *The Journal of Politics* 29, 3: 505–534.

Gomez, James, and Han Leong Chang. 2013. "New Media and General Elections: Online Citizen Journalism in Malaysia and Singapore." In *Democracy, Media, and Law in Malaysia and Singapore: A Space for Speech*, ed. Andrew T. Kenyon, Tim Majoribanks, and Amanda Whiting. Abingdon, UK: Routledge.

Greenberg, Sarah, and Thomas B. Pepinsky. 2013. "Data and Maps for the 2013 Malaysian General Elections." Department of Government Working Paper. Ithaca, NY: Cornell University.

Hron, K., P. Filzmoser, and K. Thompson. 2012. "Linear Regression with Compositional Explanatory Variables." *Journal of Applied Statistics* 39, 5: 1115–1128.

Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century.* The Julian J. Rothbaum Distinguished Lecture Series. Norman: University of Oklahoma Press.

Khoo, Boo Teik. 2013. "The 13th General Elections (GE13) in Peninsular Malaysia: An Analysis of Issues, Outcomes, and Implications." In *13th General Election in Malaysia: Issues, Outcomes, and Implications*, ed. Boo Teik Khoo. Chiba-shi, Japan: Institute of Developing Economies.

Lee, Kuan Yew. 1998. *The Singapore Story: Memoirs of Lee Kuan Yew.* Singapore: Marshall Cavendish Editions.

Leete, Richard. 1996. *Malaysia's Demographic Transition: Rapid Development, Culture, and Politics.* Kuala Lumpur: Oxford University Press.

Leong, Keith Yu Keen. 2012. *The Future of Pakatan Raykat: Lessons from Selangor.* Kuala Lumpur: Institute for Democracy and Economic Affairs.

Lerner, Daniel. 1964. *The Passing of Traditional Society: Modernizing the Middle East.* London: The Free Press of Glencoe.

Lian, Kwen Fee, and Jayanath Appudurai. 2011. "Race, Class and Politics in Peninsular Malaysia: The General Election of 2008." *Asian Studies Review* 35, 1: 63–82.

Lim, Hong Hai. 2002. "Electoral Politics in Malaysia: 'Managing' Elections in a Plural Society." In *Electoral Politics in Southeast and East Asia*, ed. Aurel Croissant and Marei John. Singapore: Friedrich-Ebert-Stiftung, Office for Regional Co-operation in Southeast Asia.

Liow, Joseph. 1999. "Crisis, Choice and Change: Malaysian Electoral Politics at the End of the 20th Century." *Asian Journal of Political Science* 7, 2: 45–74.

Liow, Joseph, and Afif Pasuni. 2010. "Debating the Conduct and Nature of Malaysian Politics: Communalism and New Media Post–March 2008." *Journal of Current Southeast Asian Affairs* 4, 4: 39–65.

Lipset, Seymour M. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *The American Political Science Review* 53, 1: 69–105.

Mansor, Mohd Noor. 1992. "The Determinants of a Malay Ethnic Alignment." PhD diss., University of Bristol.

———. 1999. "Crossing Ethnic Borders in Malaysia: Measuring the Fluidity of Ethnic Identity and Group Formation." *Akademika* 55, 1: 61–82.

McGee, Terry. 2011. "The Urbanisation Transition in Malaysia in the Twenty-First Century." *Akademika* 81, 2: 109–121.

Melson, Robert, and Howard Wolpe. 1970. "Modernization and the Politics of Communalism: A Theoretical Perspective." *American Political Science Review* 64, 4: 1112–1130.

Mohamad, Maznah. 2008. "Malaysia—Democracy and the End of Ethnic Politics?" *Australian Journal of International Affairs* 62, 4: 441–459.

Mohd Fuad, M. J., A. B. Junaidi, A. Buang, S. Selvadurai, A. C. Er, and N. Lyndon. 2011. "Ethnic Attitudes, Political Preferences, and the Politics of Stability." *World Applied Sciences Journal* 13 (Special Issue on Human Dimensions of Development): 34–38.

Moten, Abdul Rashid. 2009. "2004 and 2008 General Elections in Malaysia: Towards a Multicultural, Bi-Party Political System?" *Asian Journal of Political Science* 17, 2: 173–194.

Mutalib, Hussin. 2000. "Malaysia's 1999 General Election: Signposts to Future Politics." *Asian Journal of Political Science* 8, 1: 65–89.

Noh, Abdillah. 2014. "Malaysia 13th General Election: A Short Note on Malaysia's Continuing Battle with Ethnic Politics." *Electoral Studies* 34: 266–269.

Noor, Farish A. 2013. "The 13th Malaysian General Elections from a Sabah Perspective." *The Round Table* 102, 6: 541–548.

Norris, Pippa, and Robert Mattes. 2003. "Does Ethnicity Determine Support for the Governing Party? The Structural and Attitudinal Basis of Partisan Identification in 12 African Nations." Kennedy School of Government Research Working Papers Series. Cambridge: Harvard University.

O'Shannassy, Michael. 2009. "Beyond the Barisan Nasional? A Gramscian Perspective of the 2008 Malaysian General Election." *Contemporary Southeast Asia* 31, 1: 88–109.

Ostwald, Kai. 2013. "How to Win a Lost Election: Malapportionment and Malaysia's 2013 General Election." *The Round Table* 102, 6: 521–532.

Ottaway, Marina. 1999. "Ethnic Politics in Africa: Change and Continuity." In *State, Conflict, and Democracy in Africa*, ed. Richard Joseph, 299–318. Boulder, CO: Lynne Rienner.

Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics* 11, 6: 619–632.

Parsons, Talcott. 1975. "Some Theoretical Considerations on the Nature and Trends of Change of Ethnicity." In *Ethnicity, Theory and Experience*, ed. Nathan Glazer and Daniel Patrick Moynihan, 53–83. Cambridge: Harvard University Press.

Pepinsky, Thomas B. 2009. "The 2008 Malaysian Elections: An End to Ethnic Politics?" *Journal of East Asian Studies* 9, 1: 87–120.

Rajaratnam, Usha Devi. 2009. "Role of Traditional and Online Media in the 12th General Election, Malaysia." *Journal of the South East Asia Research Centre for Communications and Humanities* 1, 1: 35–58.

Resnick, Danielle. 2012. "Opposition Parties and the Urban Poor in African Democracies." *Comparative Political Studies* 45, 11: 1351–1378.

Saravanamuttu, Johan. 2001. "Is There a Politics of the Malaysian Middle Class?" In *Southeast Asian Middle Classes: Prospects for Social Change and Democratisation*, ed. A. Rahman Embong, 103–118. Bangi: Penerbit Universiti Kebangsaaan Malaysia.

Saw, Swee Hock. 2007. *The Population of Malaysia.* Singapore: ISEAS Publishing.

Shmuelevitz, Aryeh. 1996. "Urbanization and Voting for the Turkish Parliament." *Middle Eastern Studies* 32, 2: 162–176.

Sim, Kwang Yang. 2010. "The Cesspool of Sarawak Politics." *Hornbill Unleashed,* February 10. http://hornbillunleashed.wordpress.com/2010/12/13/12707/ (accessed October 13, 2013).

The Star. 2013. "More of a 'Malaysian and Urban' Tsunami, Says Kit Siang." May 20. www.thestar.com.my/News/Community/2013/05/20/More-of-a-Malaysian-and-urban-tsunami-says-Kit-Siang.aspx (accessed October 20, 2013).

Thompson, Eric C. 2013. "Urban Cosmopolitan Chauvinism and the Politics of Rural Identity." In *Cleavage, Connection and Conflict in Rural, Urban and Contemporary Asia*, ed. Tim Bunnell, D. Parthasarathy, and Eric C. Thompson, 161–179. Dordrecht, Netherlands: Springer.

Ufen, Andreas. 2009. "The Transformation of Political Party Opposition in Malaysia and Its Implications for the Electoral Authoritarian Regime." *Democratization* 16, 3: 604–627.

van de Walle, Nicolas. 2007. "Meet the New Boss, Same as the Old Boss? The Evolution of Political Clientelism in Africa." In *Patrons, Clients, and Policies: Patterns of Democratic Accountability and Political Competition*, ed. H. Kitschelt and S. Wilkinson, 50–67. New York: Cambridge University Press.

Weber, Max. 1978. "Urbanisation and Social Structure in the Ancient World." In *Weber: Selections in Translation*, ed. W. G. Runciman, trans. Eric Matthews. Cambridge, UK: Cambridge University Press.

Welsh, Bridget. 2013. "Malaysia's Election and Najib's Challenged Mandate." *Asia Pacific Bulletin*, 216: 1–2.

Whitby, Kenny J. 1985. "Effects of the Interaction Between Race and Urbanization on Votes of Southern Congressmen." *Legislative Studies Quarterly* 10, 4: 505–517.

Yaakob, Usman, Tarmiji Masron, and Masami Fujimaki. 2012. "Population Growth and Urbanization in Peninsular Malaysia 1911–2000." *Journal of Ritsumeikan Social Sciences and Humanities* 4: 79–101.

Yusoff, Muhammad Agus. 2001. "Sabah Politics Under Pairin." *Jati* 6: 29–48.

Journal of East Asian Studies 15 (2015), 199–226

Interpreting Ethnicity and Urbanization in Malaysia's 2013 General Election

Thomas B. Pepinsky

In this article I reinterpret Ng, Rangel, Vaithilingam, and Pillay's analysis in this issue of pro-BN voting in Peninsular Malaysia in Malaysia's 2013 general election. I show that the authors' statistical methods are inappropriate for testing whether district ethnicity predicts district-level BN vote share, and that their modeling choices result in tests of hypotheses that do not exist and cannot be derived from standard theoretical approaches to ethnic voting in Malaysia. I then provide a range of statistical evidence that supports three main conclusions: (1) ethnicity and district area (a proxy for urbanization) both predict BN vote shares at the district level, (2) neither the effect of ethnicity nor of district area can be reduced to the other, and (3) there is no interactive effect between ethnicity and urbanization. These results are in direct contradiction with the authors' results, and apply equally in Peninsular Malaysia and the entire country. I also discuss the broader issues that emerge when testing competing theories of BN vote share. KEYWORDS: *ethnicity, urbanization, elections, authoritarianism, Malaysia, statistics*

NG, RANGEL, VAITHILINGAM, AND PILLAY'S ANALYSIS (THIS ISSUE) OF ETHNICITY, urbanization, and pro-regime voting in Malaysia's 2013 general election is an important contribution to contemporary Malaysian political studies. The authors (hereafter NRVP) use advanced statistical techniques to estimate the relationships between ethnic population totals, urbanization, and constituency-level votes for the Barisan Nasional (BN) coalition in Peninsular Malaysia. By interacting a measure of ethnic composition with a measure of district area, they purport "to identify which of . . . two factors, ethnicity or urbanization, provides a stronger explanation for the erosion of the BN's popular votes in GE13 [the 2013 general election]." They conclude that "the *Chinese-Urbanization* factor is having the most dominant influence on the proportion of votes garnered by

BN," and also that "whether the constituency is an urban or rural region, an increase in the number of Bumiputera voters in that constituency, ceteris paribus, does not alter the level of support for the ruling coalition."

NRVP's article raises important questions about Malaysian politics, and the way that the authors tackle them has implications for the comparative study of ethnic politics. In the case of Malaysia, ethnicity has been the dominant framework for interpreting Malaysian politics since independence, and the durability of the BN regime has always depended on its ability to amass Bumiputera votes and, in particular, on its ability to mobilize Malay voters in Peninsular Malaysia. One consequence of the BN's strategy is that the percentage of a district's population that is Malay is a powerful predictor of the share of the vote in that district going to the BN. Recently, a wealth of qualitative data—including my own subjective impressions—suggests that urbanized Malays are no longer as closely aligned with the United Malays National Organization (UMNO) and the BN as they once were. If it could be shown that there is no longer a correlation between district-level ethnic composition and BN vote shares, and that some other factor—perhaps modernization, perhaps urbanization, perhaps some other form of social change—had replaced it, then this would be powerful evidence that the customary logic of Malaysian politics had changed in a fundamental way, with implications for the durability of the BN regime and for opposition party strategy.

This is why NRVP's analysis, which emphasizes the importance of urbanization over ethnicity, is so important to our understanding of Malaysian politics. I join NRVP in emphasizing that a comprehensive treatment of the data is necessary, but the details of that analysis matter, and unavoidably involve technical discussions of statistical specification. We must also understand the conceptual issues with "causes of effects" research designs (see Gelman 2011) that aim to adjudicate among different explanations for BN vote share. As I argue below, "horserace" approaches that pit one explanation against another by including both and their interaction in a regression model are not proper tests of competing hypotheses.

In this comment, I present a simpler analysis, one guided by the substantive problem and attentive to the complexity of making inferences from massively interactive models with highly correlated predictors. Some of the discussion below is technical in nature, but this is both unavoidable and essential to understanding how the statistical models relate to substantive questions. Taken together, the evidence supports three main conclusions.

1. Both district-level ethnic structure and district land area (a proxy for urbanization) predict BN vote shares at the district level.
2. Neither the effect of ethnicity nor that of urbanization can be reduced to the other.
3. There is no interactive effect between ethnicity and urbanization.

These results are in direct contradiction with the authors' results, and apply equally in Peninsular Malaysia and the entire country.

My analysis sounds a note of skepticism that urbanization has moderated—much less superseded—the relationship between district ethnic composition and BN vote share. Instead, it confirms that both ethnicity and urbanization are excellent predictors of BN vote share, which suggests that it would be misleading to select only ethnicity or urbanization for analysis, or to argue that only one and not the other matters. However, if we follow the authors' lead in asking which variable—ethnicity or urbanization—"provides a stronger explanation" for BN vote share, using appropriate tests for competing hypotheses, then ethnicity wins. Every model, every time.

Background

Most of the pertinent details about Malaysia's 2013 general election can be found in NRVP, so I do not repeat them again here.[1] The centerpiece of their analysis is a statistical analysis of the relationship between urbanization, ethnicity, and district-level vote returns. To my knowledge, the first peer-reviewed article in English that used regression analysis to understand ethnicity and vote returns is my own 2009 article in this journal (see Pepinsky 2009). That analysis did not consider urbanization as a competing explanation for patterns of vote returns, so it is imperative to recognize that NRVP's consideration of the competing dynamics of urbanization is an important, necessary step forward. It helps to build a more sophisticated, more nuanced characterization of district-level vote returns than one that can be achieved by looking at ethnicity in isolation.

NRVP's article—in particular, the working paper version[2]—was also part of a lively debate, during and after the election, about urbanization in Peninsular Malaysia and the declining support for the BN. Analysts in the run-up to GE13 emphasized the importance of the UMNO machine in rural areas (see, e.g., Aspinall 2013), and afterward argued that the conduct of the election and its results reflected an urban-rural divide in the Malay electorate (e.g., Aljunied 2013). Given that the BN won the election with a minority of the popular vote, emphasis naturally turned to ger-

rymandering, in particular to the rural bias in constituency delineation that tended to favor the BN (e.g., Lee 2013; Ostwald 2013). Nevertheless, there were other voices, such as Kessler (2013), who argued that

> UMNO/BN saw, as some who were not part of its campaign also understood, that the key to the election was the Malay votes. . . . It was conducted in Malay terms and directed to a Malay audience. . . . It was a campaign conducted for the votes of Malays, mainly for those of the great bulk of the more "traditionally-minded" Malays, in the Malay rural heartland areas.

But Kessler's formulation is instructive. Even after decades of urbanization, Malay voters still tend to be rural voters, and the Malay constituencies in which UMNO and the BN needed to win were therefore rural constituencies.

The observation that ethnicity and urbanization covary has profound implications for our ability to disentangle conceptually which one drives support for the BN. Whether using qualitative evidence or statistical modeling, we cannot simply look at rural areas and their tendency to vote BN and conclude that they do so because they are rural, rather than because they are predominantly Malay. This observation also helps to put GE13 in its proper historical political context, for ethnicity and urbanization covary in Malaysia for reasons that are critical for understanding Malaysian party politics—that is, the perceived social and economic hierarchy in colonial Malaya, which featured a largely (but not exclusively) urban Chinese population and a largely rural Malay population. The fact that the Malays were largely rural, and hence "backward," was considered part of the justification for why Malays needed a party like UMNO that would advocate in favor of their interests. It would not have made sense to separate UMNO's rural focus from its Malay focus, for historically they were one and the same, and one justified the other.

This dynamic has not much changed. A party campaigning for Malay votes in a rural district will need to emphasize rural issues. In rural areas, therefore, rural issues happen to also be Malay issues. This is not to ignore the other resources that UMNO and the BN have in rural areas. UMNO is a finely tuned machine with deep reach into rural communities. But of course, these are also Malay communities. We must be careful not to ignore the substantive weight of ethnicity when a party named the United Malays National Organization, founded to represent Malay interests, with a successful and widely known history of campaigning on—and governing on behalf of—Malay interests, campaigns for Malay votes in Malay areas.

Altogether, NRVP's analysis of ethnicity against urbanization is an important addition to the literature on Malaysian voting. But even if it is possible to distinguish between them statistically, in reality, ethnicity and urbanization are part of a single, larger political dynamic in Malaysian politics. With this in mind, I turn now to NRVP's statistical methods.

Statistical Issues

Two particular features of the data guide NRVP's statistical analysis. The first is the limited range of the dependent variable (*BN Vote Share*), which is the ratio of votes obtained by the BN to total votes cast. This variable may logically range from 0 (no votes to the BN) to 1 (all votes to the BN). There are two related issues here. The first is statistical: a linear regression may generate illogical predicted values of the dependent variable that lie outside of the feasible interval of [0,1]. The second is theoretical: it is reasonable to expect that the effect of an increase in Bumiputera population share is different for districts that are 20 percent Bumiputera versus 80 percent Bumiputera. NRVP confront both of these issues using a fractional logistic regression approach (Papke and Wooldridge 1996), which both accounts for the bounded nature of the dependent variable and uses the logit link function to structure the analysis around one natural form of nonlinearity in the effects of independent variables.[3]

There is no doubt that the limited range of the dependent variable could in principle affect inferences. However, I will demonstrate that simple ordinary least squares (OLS) regression performs extremely well in modeling the relationships among ethnicity, urbanization, and vote share, such that employing the fractional logit approach makes no substantive difference to the inferences we draw from the analysis. It is a nice application of generalized linear modeling, but it does not require us to rethink any conclusions that we might have drawn from a simple OLS analysis. One reason that most political scientists use OLS to model vote shares is that fractional regression methods rarely change substantive conclusions unless vote shares of zero appear frequently in the data (see, e.g., the discussion in Gardeazabal 2010).

The second troublesome feature of the data is the nature of district ethnic structure. For each district, there is a breakdown of ethnicity population shares F for each of four key ethnic categories: (F_{Bumi}, F_{Chinese}, F_{Indian}, F_{Other}). This type of data is known as compositional data (Aitchison 1986), and it raises a thorny problem for statistical analysis. Because $F_{\text{Bumi}} + F_{\text{Chinese}} + F_{\text{Indian}} + F_{\text{Other}} = 1$, it must be the case that increasing the

share of one group corresponds to a decrease in the share of at least one other group. But when we include each of the four terms as predictors in a regression-type analysis, interpreting coefficients requires a counterfactual statement of the type "an increase in F_i holding all F_{-i} constant." We thus have a contradiction, because we cannot logically increase, say, Bumiputera population share while holding other population shares constant.

NRVP confront this challenge by making a substantively important change in how they measure ethnicity. Rather than use F_i, they use the total ethnic population per district, T_i, which they estimate by multiplying F_i by the total number of voters in a district. Because the sums of the total ethnic populations are not constrained to add up to 1, T_i is free from the interpretation challenges associated with ethnic population shares.

The decision to replace F_i with T_i is driven entirely by the problems of using compositional data in regression-type analyses. NRVP note, appropriately, that standard solutions for compositional data involve complex transformations of the problematic independent variables that are both uninterpretable in substantive terms and still more confusing in interaction models. But their solution has the effect of changing the research question at hand from the analysis of the effect of *ethnic composition* to *ethnic population totals*. I am aware of no theory of why districts with higher raw numbers of Bumiputeras, Chinese, Indians, or others in a district would be more likely to vote one way or another, whereas a long line of research and even the most cursory observation of Malaysian politics over the past half century would suggest that the higher the Bumiputera population share, the higher the BN vote share. By measuring ethnic population totals rather than population shares, NRVP predict that Bukit Mertajam constituency in Penang (18.9 percent Bumiputera) would be comparable to Putrajaya (95.5 percent Bumiputera) simply because the total number of Bumiputera voters in each is approximately 15,000! As it turns out, the BN received 18.7 percent of the vote in Bukit Mertajam, and 69.3 percent of the vote in Putrajaya.

My prediction, moreover, emerges logically from a microfounded theory of ethnicity and partisanship in Malaysia. If (a) Bumiputera are more likely to vote for the BN than non-Bumiputera, then (b) ceteris paribus, the higher the proportion of voters in a district that are Bumiputera, the higher the BN vote share. The same prediction does not hold for population totals: even if (a) holds, then *it does not follow* that more voters in a district are Bumiputera, the higher the BN vote share.[4] Replacing F_i with T_i, then, results in a test of a theory that has not been

articulated, that does not accord with the realities of Malaysian politics, and cannot even be derived from assumptions about ethnicity and voting behavior at the individual level.

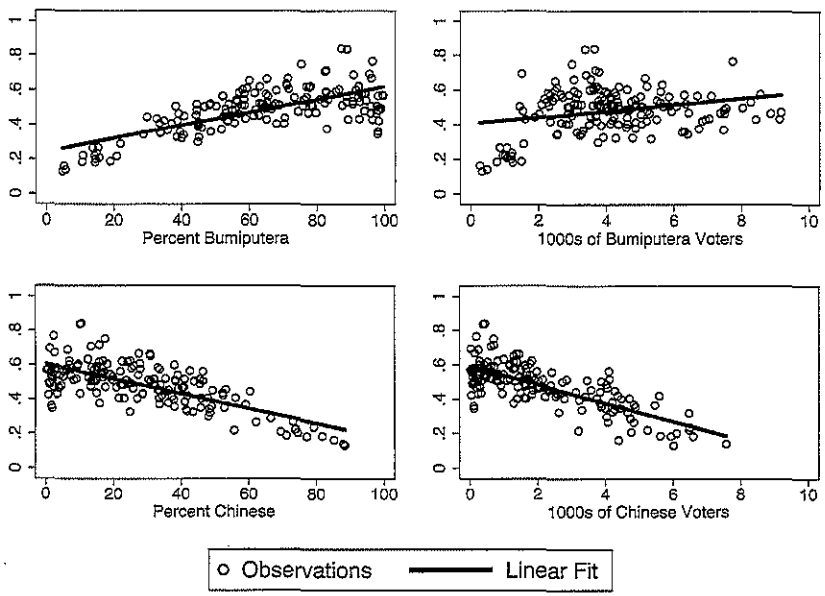
Unnoticed by NRVP is an alternative way forward. There is a simple, theoretically appropriate, and statistically sound modeling strategy for testing the effects of ethnic population shares on BN vote shares. There is no need to enter (F_{Bumi}, F_{Chinese}, F_{Indian}, F_{Other}) into the same regression. When doing so—and for now ignoring the compositional data problem—the result is a test of the effect of, for example, Bumiputera population share relative to other population share, holding Chinese and Indian population shares constant. (This is because one of the four categories will form a reference category, and will be dropped from the regression.) To test the effects of Bumiputeras relative to all others, however, we can simply enter F_{Bumi} alone into a regression. The reference category, now dropped from the analysis, will be all non-Bumiputeras (that is, Chinese, Indians, and others together). We can repeat this for each of the other three categories to produce four regressions, each of which tests whether there is a correlation between one ethnic group's population share and the percentage of votes received by the BN. Doing so preserves the substantive hypothesis about the predictive effects of ethnicity on BN votes, violates no assumptions about coefficient interpretability due to compositional data problems, and can be extended in a straightforward manner to interaction models. The cost is only several milliseconds of computing time.

Visualizing Election Results

Before showing those regression results, it is helpful to look directly at the data. In Figure 1, I plot the correlations between BN vote share and percent Bumiputera and percent Chinese (left side), and estimated number of Bumiputera and Chinese voters (right side), using NRVP's own data, which they generously shared with me.

The correlations between percentage BN vote share and percent Bumiputera and percent Chinese are strong and obvious. *No amount of statistical modeling in the rest of this comment will overturn these findings.* However, the correlations between total number of Bumiputera and BN vote share are not as strong. In fact, without the cluster of districts that have both small numbers and small proportions of Bumiputeras, total Bumiputera population would have no predictive power at all over BN vote shares. Note, however, the strong negative correlation between numbers of Chinese voters and BN vote share.

Figure 1 Ethnicity and BN Vote Shares



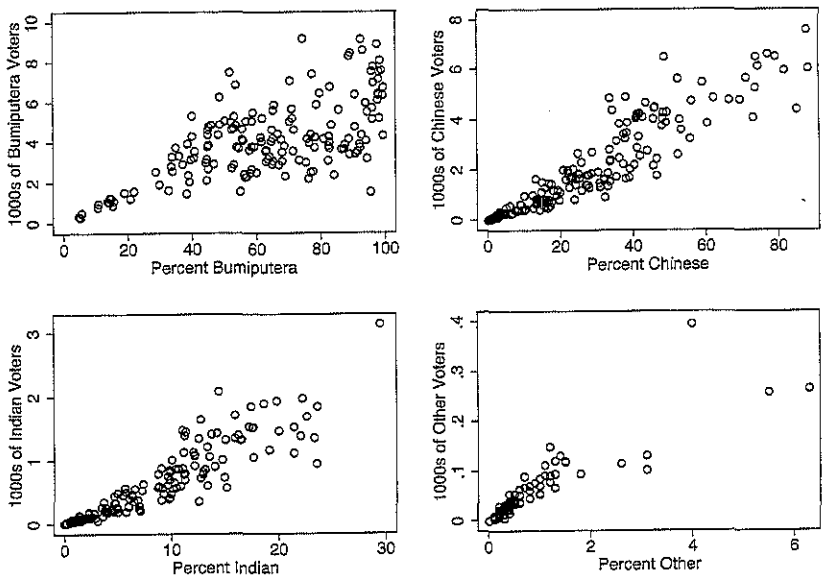
Source: Data are from NRVP.
Note: This figure displays district-level data from Peninsular Malaysian parliamentary districts that compare Bumiputera and Chinese population shares (left two plots) to BN vote shares and estimated total numbers of Bumiputera and Chinese voters to BN vote shares (right two plots).

This suggests a strong correlation between population shares and population totals for Chinese, and that is exactly what the data show. In Figure 2, I plot percentages versus population totals for all four ethnic groups.

There is always a correlation between population shares and population totals, but that in the case of Bumiputera, the variance is much larger. This has implications for statistical analysis. When predicting BN vote shares, population totals will be reasonable—albeit imperfect—proxies for the actual theoretical variable, ethnic population share. But it turns out that when using interactive multivariate models, in which eyeballing the data across multiple dimensions is not possible, imperfect proxies will generate misleading inferences.

Before proceeding to the multivariate analysis, I can also examine the relationship between population shares and urbanization. As a proxy for urbanization at the electoral district level, I use district size. It turns out that district size is highly skewed, as Figure 3 shows.

Figure 2 Percent Bumiputera Versus Total Bumiputera

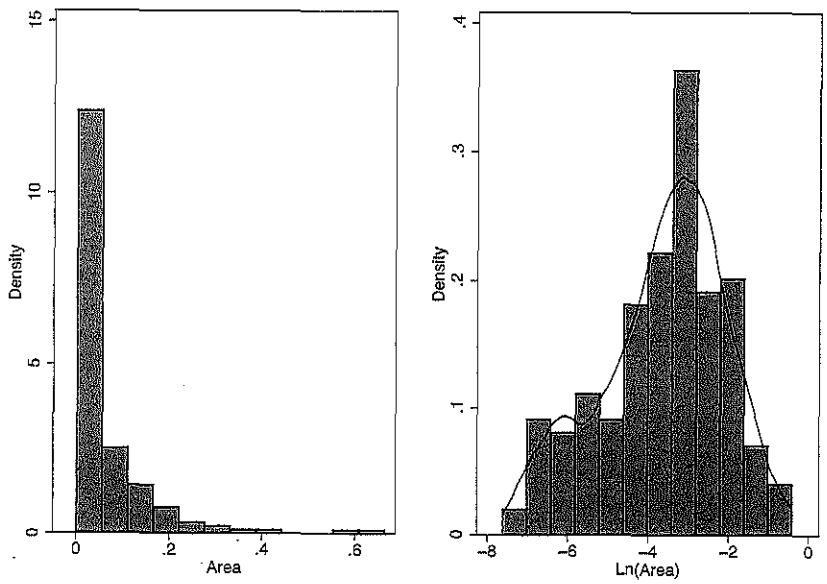


Source: Data are from NRVP.
Note: This figure displays district-level data from Peninsular Malaysian parliamentary districts that compare ethnic population shares to estimated total numbers of voters by ethnicity.

However, Figure 3 also shows that the natural logarithm of district size is closer to being normally distributed. I therefore use the natural logarithm of district size as my key measure of how urban or rural an electoral district is.

In Figure 4, I provide scatterplots of ethnic population share for Bumiputera and Chinese and the log of district area. We see that on average, larger (i.e., more rural) districts tend to be more heavily Bumiputera than smaller districts. The reverse is true for Chinese, who tend to be the predominant ethnic group in smaller, more urban districts. The correlations are not perfect, of course. If they were, it would be impossible to distinguish empirically between the effects of ethnicity and urbanization, and all comparisons of the predictive effects of ethnicity versus urbanization are identified statistically by the variation in urbanization that exists for any given ethnic structure. Yet examining the raw data in this way reveals—in a way that regression analysis cannot—that urbanization and ethnicity are highly correlated, and both predict BN vote share.

Figure 3 The Distrubution of District Area



Source: Data are from Greenberg and Pepinsky (2013).
Note: This figure displays district area and the natural logarithm of district area for Peninsular Malaysian parliamentary districts.

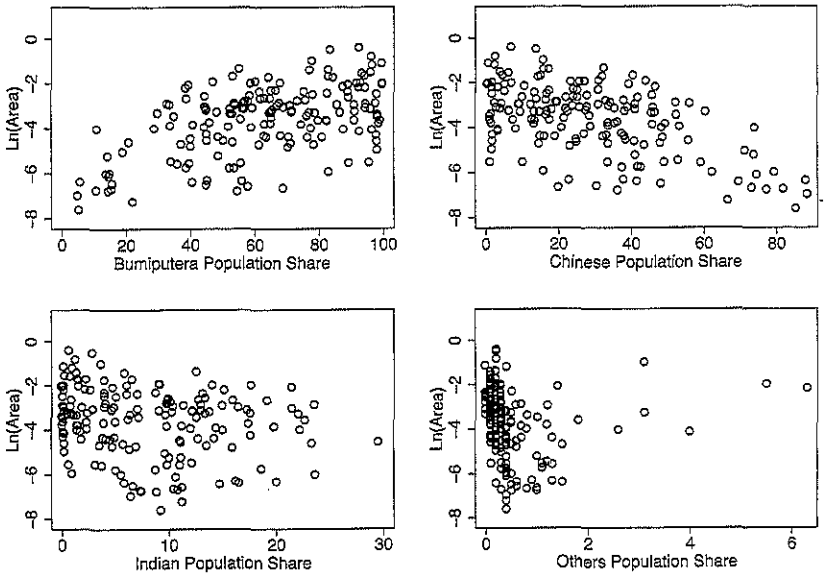
Modeling

With these visual results in hand, I turn now to a formal statistical analysis. The dependent variable is *BN Vote Share* described above. The central independent variables are *Ln(Area)* to proxy for urbanization and *% Ethnicity*$_i$ (denoted F_i above) for each of the four main ethnic groups to capture district ethnic structure. I examine a series of models that include the urbanization and ethnicity variable independently, additively, and interactively. The full model with interactions, then, is

$$BN\ Share = \beta_0 + \beta_1\ \%\ Ethnicity_i + \beta_2\ Ln(Area) + \beta_3\ \%\ Ethnicity_i \times Ln(Area) + \delta D + \varepsilon$$

Here, D is a vector of state fixed effects, and ε is an error term. I note here that I depart from NRVP by estimating robust standard errors clustered by state (rather than simple robust standard errors) throughout, although this has no substantive impact on the inferences that I draw from the results. More substantively, the state effects D capture any differences

Figure 4 Ethnic Groups by District Area



Source: Data are from NRVP and Greenberg and Pepinsky (2013).
Note: This figure displays district-level data from Peninsular Malaysian parliamentary districts that compare ethnic population shares to the natural logarithm of district size.

across states that might affect BN vote share. Given that states in the northern "Malay belt," especially Kelantan, have historically been centers of opposition to the BN, and that there is variation by state both in the distribution of district areas and of ethnic composition, including state effects will absorb any state-level factors that threaten my inferences about how ethnic structure and urbanization affect BN vote choice.

I begin by estimating models with only ethnicity and state fixed effects as the independent variables. The results appear as models 1–3 in Table 1.

As expected, ethnic population shares for Chinese and Bumiputera are excellent predictors of BN vote share. Indeed, together with state fixed effects, they alone explain most of the variation in BN vote share in Peninsular Malaysia. Results for Indian population share are markedly less strong, which is consistent with the relatively weak political position of Indian Malaysians. In model 4, I enter *Ln(Area)* as the sole predictor of BN vote share aside from the state dummies. This result too is very strong: larger (more rural) districts yield higher BN vote shares. In mod-

Table 1 Baseline Models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
% Bumiputera	0.01***				0.00***		
	(11.22)				(7.29)		
% Chinese		—0.01***				–0.00***	
		(–13.14)				(–8.40)	
% Indian			–0.01*				–0.01*
			(–2.76)				(–2.73)
Ln(Area)				0.06***	0.03***	0.02**	0.06***
				(11.22)	(4.38)	(3.51)	11.94)
N	165	165	165	165	165	165	165
Adjusted R^2	0.84	0.84	0.39	0.59	0.87	0.86	0.62
AIC	–514.84	–519.13	–297.82	–361.48	–551.12	–545.77	–375.88
BIC	–511.74	–516.02	–294.71	–358.37	–544.91	–539.56	–369.67

Notes: Each model is an ordinary least squares regression with BN vote share as the dependent variable. Each model includes state fixed effects (not reported), and standard errors are clustered by state. T-statistics in parentheses. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

els 5–7, I enter each ethnicity variable together with *Ln(Area)* to test whether the effect of one absorbs the effect of another. The results of these three models are the central findings in this analysis: the strong positive (negative) correlation between Bumiputera (Chinese) population share and BN vote share remains highly statistically significant even when controlling for district area. And the reverse is true as well, with the strong positive relationship between district area and BN vote share remaining highly statistically significant after controlling for each ethnic group's population share.

To summarize the first set of results, a simple analysis of effects of ethnicity and urbanization shows that both are excellent predictors of BN vote share, in ways that are consistent with a commonsense interpretation of Malaysian politics.

At this point the analysis might stop. However, NRVP's preferred approach to modeling the relationship between urbanization, ethnicity, and BN vote share is to interact the predictors, rather than simply entering their effects additively. Why do this? The intuition is that the effects of ethnicity might themselves depend on the level of urbanization. Uncovering these kinds of effects requires interactive models. Note, however, that the nature of the data will make it hard to test every interactive hypothesis. There are no large rural districts that are overwhelmingly Chinese, so while it is possible to calculate predicted BN vote share for a district that is both rural and overwhelmingly Chinese, such a district

does not exist (see King and Zeng 2006 for a discussion). These possibilities necessitate care in interpreting the results that we obtain from interactive models, for these calculations may be performed even if they do not make substantive sense.[5]

In Table 2, I show the results of interactive models. Models 1, 3, and 5 are identical to models 5, 6, and 7 in Table 1, and are included in Table 2 again as a reference against which to compare the interactive models.

The results are interesting. When interacting Bumiputera population share with district area, the interactive effect is miniscule and imprecisely estimated. Moreover, the standard errors on the main effect for district area rise substantially. The same nonresults for interactive effects obtain for the other two ethnic population shares, although the main effect for population size remains highly statistically significant. Yet the main effects for ethnic population share remain large and highly statistically sig-

Table 2 Interaction Models

	(1)	(2)	(3)	(4)	(5)	(6)
% Bumiputera	0.00***	0.00**				
	(7.29)	(3.80)				
% Chinese			–0.00***	–0.01***		
			(–8.40)	(–4.14)		
% Indian					–0.01*	–0.01
					(–2.73)	(–1.84)
Ln(Area)	0.03***	0.02	0.02***	0.03***	0.06**	0.07***
	(4.38)	(1.84)	(3.51)	(3.64)	(11.94)	(5.39)
% Bumiputera x *Ln(Area)*		0.00				
		(0.08)				
% Chinese x *Ln(Area)*				–0.00		
				(–1.12)		
% Indian x *Ln(Area)*						–0.00
						(–0.73)
Constant	0.35***	0.34**	0.75***	0.77***	0.78**	0.79***
	(6.31)	(3.91)	(25.37)	(20.62)	(22.30)	(17.23)
N	165	165	165	165	165	165
Adjusted R^2	0.87	0.87	0.86	0.87	0.62	0.62
AIC	–551.12	–549.15	–545.77	–548.78	–375.88	–375.90
BIC	–544.91	–539.83	–539.56	–539.47	–369.67	–366.58

Notes: Each model is an ordinary least squares regression with BN vote share as the dependent variable. Each model includes state fixed effects (not reported), and standard errors are clustered by state. T-statistics in parentheses. *$p < 0.05$; **p < 0.01; ***$p < 0.001$.

nificant. In short, these results show no evidence whatsoever of an interactive effect of ethnicity and urbanization. Viewed next to the simpler analyses in models 1, 3, and 5, it is clear that the effects of urbanization and ethnicity are better captured as additive effects.

Why are my results so different than those of NRVP? NRVP devote considerable attention to the functional form assumptions and the logical limits on the range of the dependent variable. Is it possible that my use of OLS regression explains my different results? In Table 3, I check by estimating fractional logit equivalents for every OLS model in Table 2.

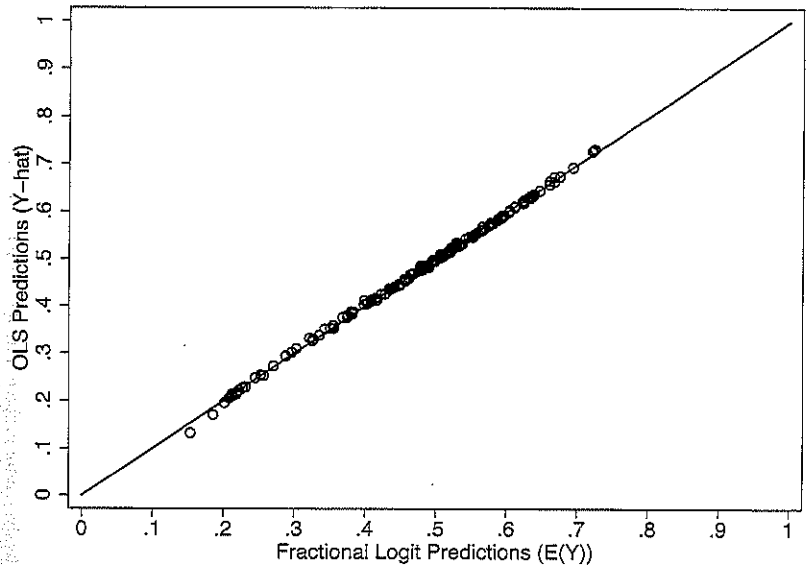
The fractional logit estimates are substantively identical to OLS estimates. We can also check to see if I obtain massively different—or illogical—predicted values from the OLS models. In Figure 5, I compare the predicted values from model 2 in Table 2 (OLS) and model 2 in Table 3 (fractional logit).

Table 3 Interaction Models, Fractional Logit Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
% Bumiputera	0.02***	0.02**				
	(6.91)	(3.32)				
% Chinese			—0.02***	–0.02***		
			(–7.93)	(–3.65)		
% Indian					–0.02**	–0.03
					(–2.79)	(–1.70)
Ln(Area)	0.11***	0.14*	0.09***	0.10***	0.24**	0.27***
	(4.38)	(2.45)	(3.70)	(2.97)	(10.91)	(5.25)
% Bumiputera x *Ln(Area)*		0.00				
		(–0.58)				
% Chinese x *Ln(Area)*				–0.00		
				(–0.36)		
% Indian x *Ln(Area)*						–0.00
						(–0.62)
Constant	–0.66**	–0.49	1.06***	1.09***	1.17***	1.22***
	(–2.77)	(–1.30)	(9.27)	(6.69)	(7.35)	(6.26)
N	165	165	165	165	165	165
AIC	147.93	149.92	147.90	149.90	150.69	152.65
BIC	154.15	159.24	154.12	159.22	156.90	161.97

Notes: Each model is an ordinary least squares regression with BN vote share as the dependent variable. Each model includes state fixed effects (not reported), and standard errors are clustered by state. T-statistics in parentheses. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Figure 5 Comparing Predictions from OLS and Fractional Logit



Notes: This figure compares OLS predicted values from model 2 in Table 2 to fractional logit expected values from model 2 in Table 3. The forty-five-degree reference line represents the point of equivalence between the two. The figure demonstrates that OLS and fractional logit predictions are nearly identical for nearly every district, and no OLS predicted value lies beyond the logical interval of [0,1].

The predictions are essentially the same, and no OLS predicted values are anywhere close to 0 or 1. There are no grounds to worry that the functional form assumptions of OLS are generating faulty inferences.

Could it be that I have misinterpreted the results by focusing on regression coefficients? Brambor, Clark, and Golder (2006) remind us that coefficients and standard errors in tabular regression outputs are not easy to interpret. So in Figure 6, I plot both expected values and marginal effects from models 2 and 4 in Table 3, alongside their 95 percent confidence intervals.

Look first at the top two plots. The top left figure plots the predicted BN vote share across the range of values of *Ln(Area)* for different levels of Bumiputera population share. Consistent with the interpretation above, the larger the area, the higher the predicted BN vote share—this is what the upward-sloping lines convey. Furthermore, the higher the Bumiputera population share, the higher the predicted BN population share—this

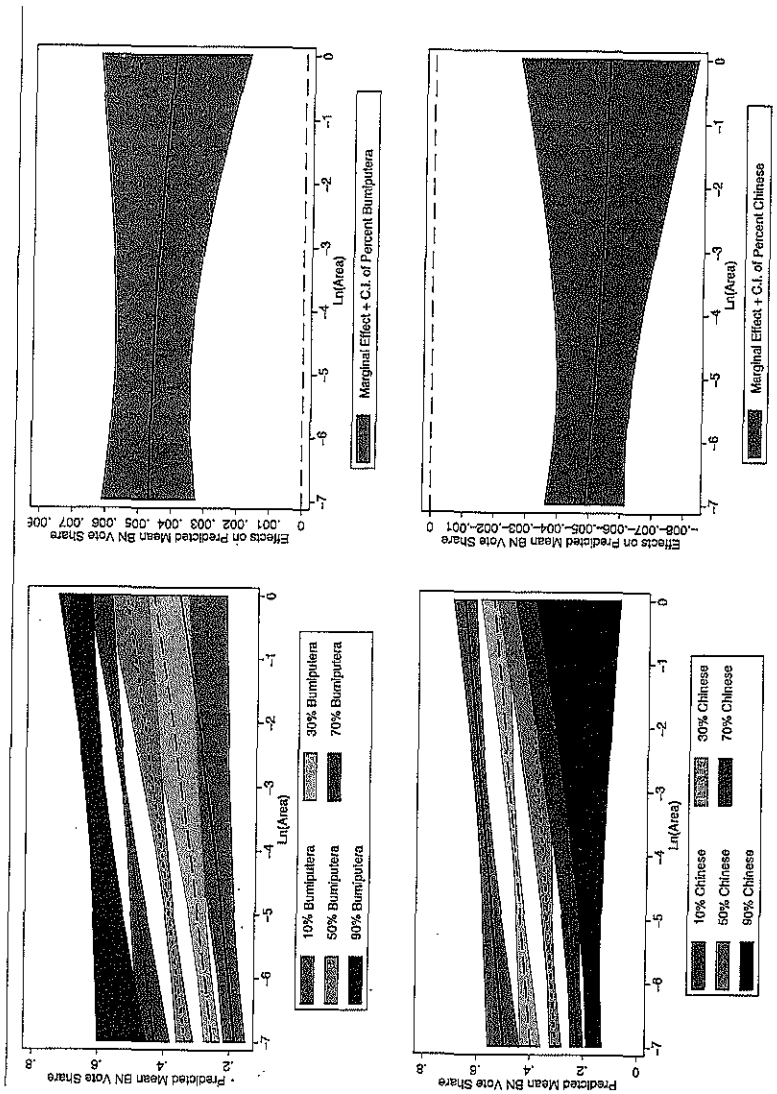


Figure 6 Predicted Values and Marginal Effects

Notes: These figures display predicted BN vote shares by district for different Bumiputera and Chinese population shares (left two plots) and the marginal effects of Bumiputera and Chinese population shares (right two plots). Both predicted vote shares and marginal effects are calculated across the range of values of *Ln(Area)*. The predictions were derived from models 2 and 4 in Table 3.

is what the five separate shaded regions show. More important, the five lines all rise in parallel, which indicates that the effect of urbanization is roughly the same regardless of the value of Bumiputera population share. This conclusion can also be drawn from the top right plot, which shows the marginal effect of an increase in Bumiputera population share across levels of *Ln(Area)*. The line slopes downward a bit, but the range of the predicted marginal effects is always far smaller than the 95 percent confidence band. And the marginal effect of Bumiputera population share is always positive. There is no evidence that the effects of Bumiputera population share depend in any way on district size.

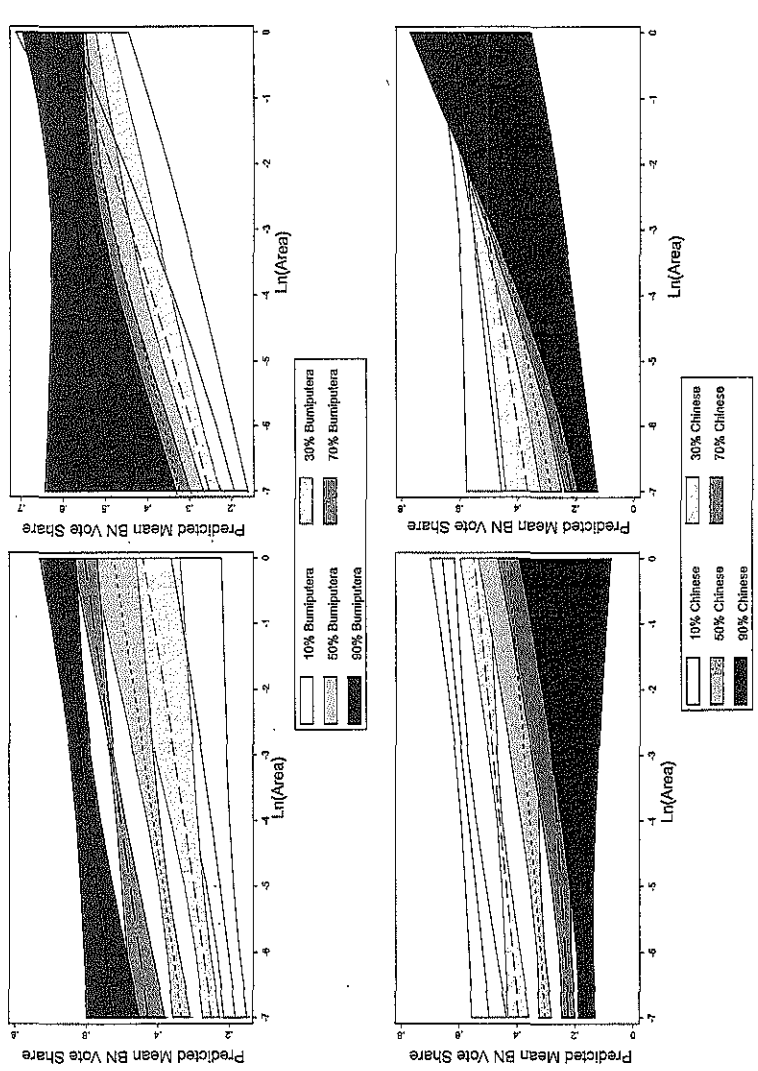
The results for Chinese population share are exactly the reverse. The higher the Chinese population share, the lower the predicted BN vote share, even allowing for the finding that the larger the district area, the higher the predicted BN vote share. Moreover, the marginal effect of Chinese population share is always negative, and while the magnitude increases slightly in larger districts, the range of the predicted marginal effects always lies well within the 95 percent confidence band. Note further the wide confidence intervals around the darkest line, corresponding to the predicted BN vote shares for a 90 percent Chinese district, in large districts. This reminds us that any predictions about the effects of Chinese ethnicity in rural districts should be treated with caution. In sum, the findings from Figure 6 demonstrate once again that both ethnicity and urbanization are strong predictors of vote share, and that there is no evidence of any interactions between the two.

If neither functional form assumptions nor interpretation issues explain the difference between my results and those of NRVP, what does? There are two answers: my use of a more theoretically appropriate and substantively interpretable measure of ethnicity,[6] and my inclusion of state fixed effects *D*. I have already shown that ethnic population shares are more appropriate than ethnic totals, but before proceeding I discuss the importance of accounting for state-specific effects.

State fixed effects have important consequences for how we interpret the interactive effects of ethnicity and urbanization. In Figure 7, I compare the predicted BN vote shares from models 2 and 4 in Table 3 with the same results obtained from fractional logit models without state effects.

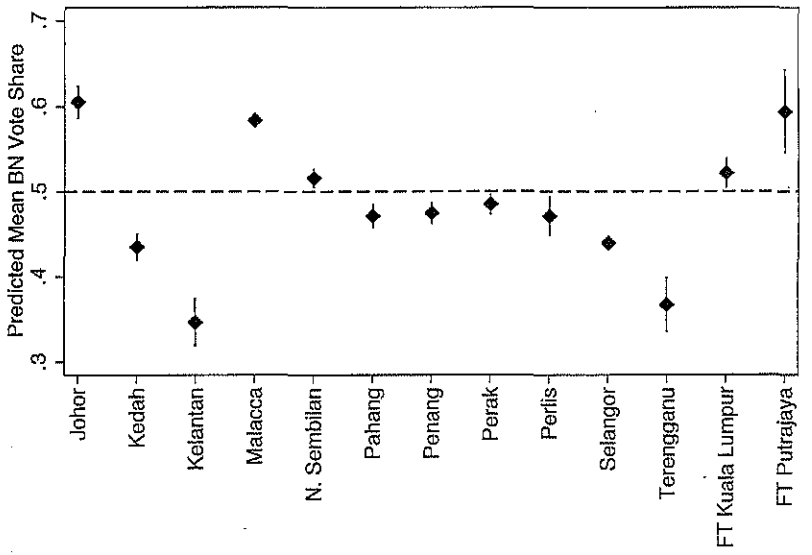
The differences between fixed effects and nonfixed effects models are quite apparent. The effects of ethnicity on BN vote share disappear in larger districts when we ignore state fixed effects, and, furthermore, there is no evidence of an effect of Bumiputera population share on BN vote share for any level of urbanization. Such results might be interpreted

Figure 7 Interactive Results, with or Without State Effects



Notes: These figures display predicted BN vote shares by district for different Bumiputera and Chinese population shares across the range of values of *Ln(Area)* for fractional logit models including fixed effects (models 2 and 4 in Table 3, left two plots) and otherwise identical models without fixed effects (right two plots).

Figure 8 Heterogeneity in BN Support by State (Peninsular Malaysia)



Notes: This figure plots predicted BN vote share across states, net of the effects of Bumiputera population share, district area, and their interaction. The predictions were derived from model 2 in Table 3.

as evidence that urbanization matters and ethnicity only affects BN vote share among ethnic Chinese in urban areas, which is broadly consistent with NRVP's results.

However, ignoring state effects deliberately obscures the obvious variation across Peninsular Malaysia in support for the BN. The predicted BN vote shares in 2013 differ dramatically across states, as shown in Figure 8.

And because states differ in their ethnic compositions, we risk attributing the effects of state-specific histories and political conditions to our observed theoretical variables. Large rural districts in Kelantan and Terengganu differ from large rural districts in other states, even if they are all heavily Bumiputera, and accounting for these state-level differences enables a more precise analysis of how ethnicity and urbanization shape BN vote shares.

Horseracing

The analyses shown thus far demonstrate that ethnicity and urbanization both predict vote choice extremely well.[7] This has an effects of

causes approach rather than a causes of effects approach (see Gelman 2011), for I have only sought thus far to characterize the predictive power of ethnicity and district area, not to select a cause of the distribution of BN vote shares across Peninsular Malaysian districts. Yet NRVP have a different aim: "the aim of this study is to identify which of the two factors, ethnicity or urbanization, provides a stronger explanation for the erosion of BN's popular votes in GE13." Theirs is a causes of effects approach.

I am sympathetic to NRVP's interest in knowing whether urbanization or ethnicity is a stronger explanation for why Malaysian electoral returns are the way that they are. My personal view, as an observer of Malaysian politics, is that ethnicity is an essential, fundamental factor in Malaysian politics. Yet realism tempers my sympathy for their instinct to view ethnicity and urbanization as competing explanations for Malaysian politics. There is no objective reason to believe that either ethnicity or urbanization is *the* essential driver of Malaysian politics. Instead, I suspect that the instinct to look for effects of urbanization that supersede those of ethnicity is driven by the hope among many Malaysians and political observers for a shift toward a postethnic Malaysian politics, and the belief that statistical analysis of the electoral results might provide evidence that this has taken place.[8]

For an effects of causes research design, multiple regression—when viewed as a way to illustrate causal relationships instead of just as a way to summarize partial correlations—assumes that one set of outcomes can have multiple causes. There is much less agreement about how to formally compare or adjudicate among different causes of effects. For some, the entire endeavor is ill-posed: what does it mean to assert that some explanation is "the cause of" some effect (Gelman and Imbens 2013)? One way to do this is to compare the extent to which two independent variables explain the variation in a dependent variable—in this case, do rural/urban differences explain more about the electoral results than ethnicity does? Unfortunately, in the present application, both explain a lot of variation in BN vote shares.

There are various other kinds of model selection procedures that can be used to select which model does "better" according to some metric, such as comparing R^2 as a measure of fit, comparing Akaike and Bayes information criteria, and the J and Cox-Pesaran tests. Recently, Imai and Tingley (2012) provided a very different way to think about this problem. We have two theories of what determines BN votes at the district level: ethnicity and urbanization. These two theories imply two different hypotheses. The hypotheses are non-nested: ethnicity is not a subset of urbanization, nor the other way around. Imai and Tingley propose that we can compare any set of theories using finite mixture models to compare the proportion of the cases being analyzed that are "statistically significantly consistent" with one theory versus the other.

So despite my own belief that both ethnicity and urbanization are good explanations for BN vote shares in Peninsular Malaysia, it is possible to follow NRVP, assume that explanations based on ethnicity and urbanization really are mutually exclusive explanations for BN vote share, and then consider the various methods for adjudicating between them. To repeat, this assumption that the two theories compete with one another is a theoretical assumption rather than an empirical result—it also ignores the more comprehensive additive or interactive models—yet in what follows, I proceed under this maintained assumption to see what happens. Unlike NRVP, though, my strategy does not rely on interaction terms,[9] but instead draws on established approaches to model selection and the testing of non-nested hypotheses.

The very simplest way to compare models is to compare the adjusted R^2, or the percentage of the total variation in the dependent variable that is explained by the independent variables (with a penalty applied for complex models that might overfit the data). It is worth pausing to emphasize that comparing R^2 is *very bad statistical practice*, especially from an effects of causes perspective. However, if we interpret the task of comparing theories as measuring the proportion of variance in BN vote shares explained by the different models, adjusted R^2 does this (King 1986, 677–678). We see that in Table 1, adjusted R^2 is higher for model 1 and model 2 (ethnicity) than for model 4 (district area). In a head-to-head contest between ethnicity and urbanization, score one for ethnicity.

More sophisticated model selection procedures for non-nested hypotheses include comparisons of Information Criteria, the J test, and the Cox-Pesaran test. The Akaike Information Criterion (AIC) and the Bayes Information Criterion (BIC) are lower in models 1 and 2 and 4. Score one more for ethnicity. The J test and Cox-Pesaran tests, interestingly, are uninformative because *each test rejects both models*.[10] This can happen when both models fit the data well, as is the case here. While this is not a victory for ethnicity over urbanization per se, it does raise another red flag about the wisdom of conceiving of these two theories as mutually exclusive.

Finally, consider the mixture modeling approach proposed by Imai and Tingley. Table 4 displays two quantities from each of two mixture models, one using Bumiputera population share and district area (equiv-

Table 4 Mixture Model Results

Model	Prior Probability	Number of Observations
Model 1 (Bumiputera)	0.871	143
Model 4 (*Ln(Area)*)	0.129	22
Model 2 (Chinese)	0.854	144
Model 4 (*Ln(Area)*)	0.146	21

Notes: The second column displays the mean of the estimated prior probabilities that each observation is consistent with each model. The third column displays the number of observations that are statistically significantly consistent with each model.

alent to comparing model 1 with model 4 from Table 1), the other using Chinese population share and district area (equivalent to comparing model 2 with model 4).

The second column displays the mean of the estimated prior probabilities that each observation is consistent with models 1, 2, or 4. The third column displays the number of observations that are statistically significantly consistent with model 1, 2, or 4. Together, the results are unambiguous evidence that more district election results are consistent with an explanation based on ethnicity than one based on urbanization. Score these results as the final piece of evidence in favor of ethnicity over urbanization.

I conclude this discussion by emphasizing one more time that *every piece of data that we have* indicates that it is misleading to ask whether *either* ethnicity *or* urbanization explains BN vote shares in Peninsular Malaysia: not just the results from multivariate analyses, which show that both are strong predictors even when in the same model, or a historical perspective that shows how the two variables are conceptually linked, but also additional statistical results comparing multivariate models to the single-explanation models. Additive and interactive models of BN vote share have higher adjusted R^2 and lower AIC and BIC scores than either single explanation model (see the last rows in Table 1 and Table 2). Likelihood ratio tests easily reject both individual models in favor of the additive model (they also fail to reject the additive model in favor of the interactive model). The mixture modeling approach overwhelmingly selects the additive model over either individual model (and also over the interactive model).[11] These results are strong evidence that both ethnicity and urbanization matter; the effects of neither urbanization nor ethnicity can be reduced to the other.

Extending the Analysis Throughout Malaysia

Finally, I extend this analysis to cover all of Malaysia, including the states of Sabah and Sarawak and the Federal Territory of Labuan in East Malaysia in addition to Peninsular Malaysia. To do this, I augment the data on Bumiputera and Chinese population shares and BN vote share from NVRP with data scraped from the website http://undi.info in 2013 (Greenberg and Pepinsky 2013). I then rerun the previous analyses, presenting the key results in Table 5 and Figure 9.

Begin first with Table 5. Comparing models 1 and 2 (Peninsular Malaysia only) to models 3 and 4 (all of Malaysia, identical to models 5 and 6 in Table 1) reveals that Bumiputera and Chinese population shares continue to be strong predictors of BN vote share, net of state effects, when we expand the sample to include all of Malaysia. However, the same is not true for *Ln(Area)*, where the coefficient estimate is not significant at conventional levels. Models 5 and 6 confirm that the same result holds when using fractional logit instead of OLS.

Table 5 Results for All of Malaysia

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Bumiputera	0.00***		0.00***		0.02***		0.02***	
	(7.29)		(9.28)		(8.92)		(5.52)	
% Chinese		−0.00***		−0.01***		−0.02***		−0.02***
		(−8.40)		(−11.42)		(−11.04)		(−6.50)
Ln(Area)	0.03***	0.02**	0.01	0.01	0.05	0.04	0.13***	0.03
	(4.38)	(3.51)	(1.42)	(1.17)	(1.50)	(1.19)	(3.48)	(0.60)
% Bumiputera x *Ln(Area)*							−0.00	
							(−1.82)	
% Chinese x *Ln(Area)*								0.00
								(0.93)
Constant	0.35***	0.75***	0.26**	0.67***	−1.04***	0.73***	−0.69*	0.71***
	(6.31)	(25.37)	(3.57)	(16.02)	(−3.38)	(4.03)	(−2.43)	(3.42)
N	165	165	222	222	222	222	222	222
Adjusted R^2	0.87	0.86	0.82	0.82				
AIC	−551.12	−545.77	−595.56	−593.19	195.97	195.94	197.89	197.92
BIC	−544.91	−539.56	−588.76	−586.38	202.77	202.74	208.10	208.13

Notes: This model compares results for Peninsular Malaysia only (models 1 and 2) with results from all of Malaysia (models 3–8). Each model uses BN vote share as the dependent variable. Models 1–6 are ordinary least squares regressions, and models 7 and 8 are fractional logit regressions. Each model includes state fixed effects (not reported), and standard errors are clustered by state. T-statistics in parentheses. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

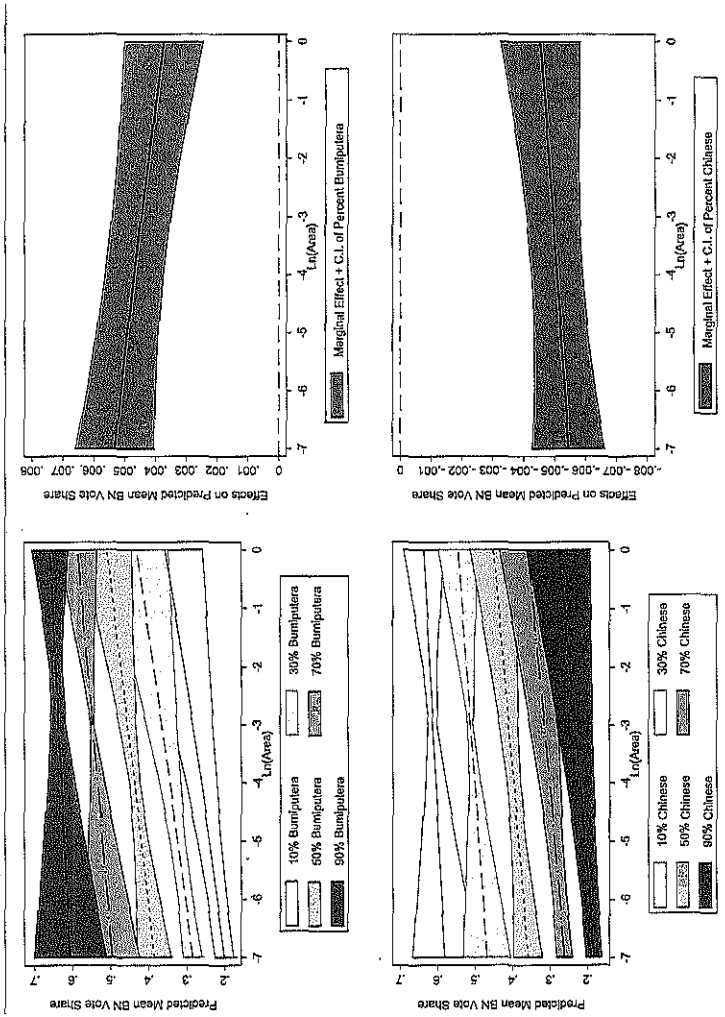


Figure 9 Interactive Results, All of Malaysia

Notes: These figures display predicted BN vote shares by district for different Bumiputera and Chinese population shares (left two plots) and the marginal effects of Bumiputera and Chinese population shares (right two plots). Both predicted vote shares and marginal effects are calculated across the range of values of *Ln(Area)*. The predictions were derived from models 7 and 8 in Table 3, and cover all 222 parliamentary districts in Malaysia.

Models 7 and 8 test the interactive hypotheses, with predicted values and marginal effects displayed in Figure 9. Interestingly, it is only in *these* models where we uncover limited evidence of an interactive effect of urbanization and ethnicity. Specifically, the top right panel demonstrates that while the marginal effect of Bumiputera population share on BN vote share is always positive and statistically significant, there is evidence that the magnitude of this effect decreases when comparing the smallest to the largest districts. This difference is statistically significant at the $p < 0.1$ level. Of course, this interaction does not eliminate the predictive effects of ethnicity on vote share, but it does modestly attenuate the size of that effect in the largest districts.

Conclusion

This article has shown that NRVP's substantive conclusions about ethnicity and urbanization are incorrect, driven by statistical modeling choices that are not appropriate for analyzing the additive and interactive effects of the two explanations for district vote returns. A simpler yet more theoretically precise statistical analysis yields a wealth of findings, but together they point to three conclusions: (1) ethnicity and urbanization both predict BN vote shares at the district level, (2) neither the predictive effects of ethnicity nor those of urbanization can be reduced to the other, and (3) there is no evidence of an interactive effect between ethnicity and urbanization. These results hold both for Peninsular Malaysia and the entire country.

Thomas Pepinsky is associate professor in the Department of Government and associate director of the Modern Indonesia Project at Cornell University. He is author of *Economic Crises and the Breakdown of Authoritarian Regimes: Indonesia and Malaysia in Comparative Perspective* (2009), as well as articles in the *American Journal of Political Science, British Journal of Political Science, Economics and Politics, International Studies Quarterly, World Development, World Politics*, and other venues.

Notes

1. This section draws on an earlier post on my blog, http://tompepinsky.com/2013/05/16/rural-or-malay-contending-perspectives-on-ge13-1/.

2. That working paper version is available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2395091. Its conclusions were more pointed than the

current version. It argued that "for any given parliamentary constituency classified as either rural, semi-urban or urban, voters have a similar voting pattern regardless of ethnicity. Therefore, the differences in the voting patterns for BN stems from the urbanisation factor instead" (p. 16).

3. The logit link function imposes a particular nonlinear functional form on the effects of predictor variables. Some readers may not be aware that it, too, is an assumption like any other, made for convenience and interpretability rather than explicitly grounded in a theory. Thus NRVP's observation that an OLS regression assumes linear effects is true, but it is not an argument *tout court* against using OLS rather than fractional logit, which replaces this linearity assumption with a different assumption about the form that nonlinearity takes. See Aldrich and Nelson (1988, 24–37) for a full discussion.

4. It is also not the case that (b) logically entails (a). It is possible that districts with higher Bumiputera population shares have higher BN vote shares for reasons other than a pro-BN bias among Bumiputeras. It could be, for example, that non-Bumiputera voters unanimously vote for the BN *only if* they are small minorities. Or it could be that Bumiputeras happen to live in rural areas, and rural voters vote for the BN. The district-level aggregate patterns cannot resolve these competing theories. This problem of uncovering individual behavior from collective behavior is known as the ecological inference problem, and has been the subject of intense study for decades (Kousser 2001). For one provisional attempt to solve the ecological inference problem in the context of Malaysia's 2008 election, see Pepinsky (2009).

5. Of course, the same is true for additive models as well, but the subtleties of interpreting interactive models appear to generate particular challenges in interpretation.

6. One might still wonder about the correlations between district population totals (which is one component of NRVP's measures of ethnic population totals) and BN vote share. In separate results, available upon request, I can demonstrate that accounting for district population total (either alone or in a triple interaction with both ethnicity and district area) has no substantive consequences for inferences about ethnicity and urbanization.

7. This section draws on an earlier post on my blog, http://tompepinsky.com/2013/05/18/rural-or-malay-contending-perspectives-on-ge13-2/.

8. Eric Thompson (2013) uses the term "urban chauvinism" to describe some of the interpretations of the results of GE13 that emphasize an urban-rural divide. I highlight this here as a reminder that nonethnic explanations for GE13 results are no less subject to normative biases than are explanations that highlight patterns in district ethnicity and BN vote share.

9. Indeed, while NRVP explicitly state that they wish to "identify which of the two factors, ethnicity or urbanization, provides a stronger explanation for the erosion of BN's popular votes in GE13," it is not immediately clear how any of their statistical analyses actually answer that question.

10. Results are available from the author upon request.

11. Results for mixture models and likelihood ratio tests are available upon request from the author.

References

Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*. New York: Chapman and Hall.

Aldrich, John H., and Forrest D. Nelson. 1988. *Linear Probability, Logit, and Probit Models*. Newbury Park: Sage Publications.

Aljunied, Khairun. 2013. "How Malays Voted at GE13." New Mandala, May 9. http://asiapacific.anu.edu.au/newmandala/2013/05/09/how-malays-voted-at-ge13.

Aspinall, Edward. 2013. "Triumph of the Machine." Inside Story, May 9. http://insidestory.org.au/triumph-of-the-machine/.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14, 1: 63–82.

Gardeazabal, Javier. 2010. "Vote Shares in Spanish General Elections as a Fractional Response to the Economy and Conflict." Economics of Security Working Paper 33. www.diw.de/documents/publikationen/73/diw_01.c.356877.de /diw_econsec0033.pdf.

Gelman, Andrew. 2011. "Causality and Statistical Learning." *American Journal of Sociology* 117, 3: 955–966.

Gelman, Andrew, and Guido Imbens. 2013. "Why Ask Why? Forward Causal Inference and Reverse Causal Questions." National Bureau of Economic Research (NBER) Working Paper 19614.

Greenberg, Sarah, and Thomas B. Pepinsky. 2013. "Data and Maps for the 2013 Malaysian General Elections." Working Paper, Department of Government, Cornell University.

Imai, Kosuke, and Dustin Tingley. 2012. "A Statistical Method for Empirical Testing of Competing Theories." *American Journal of Political Science* 56, 1: 218–236.

Kessler, Clive J. 2013. "Malaysia's GE13: What Happened, What Now? (Part 1)." New Mandala, June 12. http://asiapacific.anu.edu.au/newmandala/2013/06/12/malaysias-ge13-what-happened-what-now-part-1/.

King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30, 3: 666–687.

King, Gary, and Langche Zeng. 2006. "The Dangers of Extreme Counterfactuals." *Political Analysis* 14, 2: 131–159.

Kousser, J. Morgan. 2001. "Ecological Inference from Goodman to King." *Historical Methods* 34, 3: 101–126.

Lee, Hock Guan. 2013. "Steadily Amplified Rural Votes Decide Malaysian Elections." *ISEAS Perspective*, no. 34.

Ng, Jason Wei Jian, Gary John Rangel, Santha Vaithilingam, and Subramaniam S. Pillay. "2013 Malaysian Elections: Ethnic Politics or Urban Wave?" *Journal of East Asian Studies* (this volume).

Ostwald, Kai. 2013. "How to Win a Lost Election: Malapportionment and Malaysia's 2013 General Election." *The Round Table* 102, 6: 521–532.

Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics* 11, 6: 619–632.

Pepinsky, Thomas B. 2009. "The 2008 Malaysian Elections: An End to Ethnic Politics?" *Journal of East Asian Studies* 9, 1: 87–120.

Thompson, Eric C. 2013. "GE13 and the Politics of Urban Chauvinism." New Mandala, May 14. http://asiapacific.anu.edu.au/newmandala/2013/05/14/ge13-and-the-politics-of-urban-chauvinism.

Journal of East Asian Studies 15 (2015), 227–241

Rejoinder: The Authors Respond to "Interpreting Ethnicity and Urbanization in Malaysia's 2013 General Election"

Jason Wei Jian Ng, Gary John Rangel, Santha Vaithilingam, and Subramaniam S. Pillay

In this article we respond to Thomas Pepinsky's commentary on our article "2013 Malaysian Elections: Ethnic Politics or Urban Wave?" (both in this issue). We confirm that both ethnicity and urbanization play important roles in determining the incumbent ruling party's percentage vote share in the thirteenth general election. In doing so, we address the various econometric issues raised by Pepinsky and clearly explain the advantages of our econometric methodology vis-à-vis the OLS analysis espoused by Pepinsky. Our main results indicate that Barisan National's (BN) vote share from Bumiputera voters, regardless of urbanization levels of the parliamentary constituency, is below the 50 percent threshold. This result is surprisingly compensated by the more than 50 percent support for BN when Chinese voters are a small minority of the electorate. We also argue that Pepinsky's statement that Malay voters are predominantly rural voters is inaccurate and provide evidence to the contrary. KEYWORDS: *Malaysia, thirteenth general election, ethnic politics, fractional logit response model, urbanization, rural*

PEPINSKY'S COMMENTS AND THE UNDERLYING ARTICLES ON WHICH THEY ARE based have become part of an important debate on the role of ethnicity in Malaysian elections. The issues raised are both substantive and technical. In this response, we revisit and reconfirm our core empirical findings that Bumiputera support for Barisan National (BN) is below 50 percent regardless of parliamentary seat classification, and Chinese voters' behavior is surprisingly not homogeneous but dependent on whether they are small in numbers or make up a substantial proportion of the electorate in a particular seat. Chinese voters help tilt the balance in BN's favor, especially in rural areas where they make up a small minority of the electorate.

On the technical front, we do not attempt to answer every issue that Pepinsky has raised, but we have done additional analyses to show why the fractional logit methodology is superior to the OLS method. Pepinsky also commented that our use of the ethnic population total has the effect of changing the research question at hand from the analysis of the effect of *ethnic composition* to *ethnic population totals*. However, we argue that the use of ethnic population totals, *in the context of our model specification,* allows the interpretation of the results to be in terms of relatives (or proportions). While it may not be the best way to model ethnicity, it is a better option to model the data. In addition, we will demonstrate how the proportion of ethnic voters as used in Pepinsky's model is not able to identify the subtleties in the results, unlike our model, which uses total ethnic population. We follow up with an in-depth explanation of the key contributions of our article and address some of the claims made by Pepinsky.

Summary of Results and Methodology

Our aim in the article published in this volume on the 2013 Malaysian general election (GE13) as to identify which of the two factors, ethnicity or urbanization, provides a stronger explanation for the erosion of BN's popular vote. We do not assume that ethnicity and urbanization are mutually exclusive, as argued by Pepinsky, but instead, our analysis allows for the interaction of both factors. Our findings suggest that although the results of Malaysia's GE13 displayed an ethnic effect, complementing Pepinsky's main finding in his commentary, rapid urbanization of the country also played a role in determining the outcome of the election. Malaysians across ethnic lines voted overwhelmingly for Pakatan Rakyat (PR) in urban areas while support for BN remained reasonably strong in rural areas. Our results therefore do not rule out the ethnicity effect. Pepinsky finds that "1. Both district-level ethnic structure and district land area (a proxy for urbanization) predict BN vote shares at the district level. 2. Neither the effect of ethnicity nor that of urbanization can be reduced to the other. 3. There is no interactive effect between ethnicity and urbanization."

In short, the differences in results can be attributed to differences in the econometric modeling of the data, namely, model specification and choice of scaling of ethnic variables. Table 1 summarizes the differences in methodology between the two studies.

When confronted with different models reporting starkly contradicting results, how should one determine which of the two models is "correct"? It is helpful to note the popularly cited maxim, "Essentially, all models are wrong, but some are useful" (Box and Draper 1987, 424). All models are wrong largely because no one ever knows the true model specification. The specified econometric model is a simplified representation of reality. Therefore, the practical question to ask is not which model is wrong, but rather, "how wrong do [the models] have to be to not be useful?" (Box and Draper 1987, 74). Against this backdrop, we use this guiding principle to further justify the choice of our model specification and variables.

Table 1 Differences in Econometric Methodology Between Ng et al. (2015) and Pepinsky (2015)

	Ng, Rangel, Vaithlilingam, Pillay	Pepinsky
Econometric model	Fractional response model	Linear OLS
Independent variables	1. Ethnic population totals 2. Area 3. Interaction terms between ethnic population total and Area	1. Ethnic composition 2. Area 3. State fixed effects
Model specification	A single econometric model containing all four ethnic population totals, Area, and interaction terms	Several econometric models with each model containing only one ethnic population composition, state fixed effects, subsequently augmented by Area

Fractional Logit or Linear OLS Model?

Pepinsky casts doubt on the appropriateness and effectiveness of the use of a fractional response logit model in our article, demonstrating that "simple ordinary least squares (OLS) regression performs extremely well in modeling the relationships between ethnicity, urbanization, and vote share, such that employing the fractional logit approach makes no substantive difference to the inferences we draw from the analysis." Pepinsky subsequently performs empirical analysis to show that the predicted vote share, based on his model specifications, is almost identical for both OLS and the fractional logit approach.

The argument above that Pepinsky makes against our use of the fractional logit model is an example of the difference in disposition between taking either a theory-driven or data-driven approach. While largely similar, econometrics is predominantly theory driven while statistics tend to

be data driven. Therefore, an econometrician develops a model based on economic (and other relevant) theories while a statistician may build a model after looking at datasets. The econometrician subsequently confronts the model with datasets to test the theory. The interested reader can refer to Rob Hyndman's blog post[1] for interesting insights into the differences between the two. In this context, it can be said that our econometric model is theory driven while Pepinsky's model is data driven.

It should be noted that Pepinsky makes the observation that the OLS model performs as well as the fractional logit model ex post (i.e., after the data have been observed and modeled). On the contrary, the choice of the fractional logit model is theory driven—the dependent variable to be modeled (vote share) is a proportion quantity known to be restricted to an interval between 0 and 1. In other words, the choice of our model was dependent on the known *nature* of the data, and not based on what the data reveal. It is along this line of reasoning that probit (and logit) models were developed to model binary dependent variables (i.e., variables that take on values of either 0 or 1), and tobit models were developed to model corner solution dependent variables (i.e., variables that have a population distribution that is spread out over a large range of positive values, but has a pileup at the value 0). These models were developed to account for the theoretical nature of the data.

From a theoretical perspective, the econometric model specified to model the proportion of vote share to BN must account for all possibilities, including the possibility of observing either a 0 or 1. However, in the event that the dependent variable does not realize values of either 0 or 1, as per the dataset observed for Malaysia's general election, the fractional logit model specified assigns 0 weights to the probability of observing these two value bounds in the log-likelihood function. Furthermore, by specifying a fractional logit model, we have not made any a priori assumptions on the restricted range of values that the dependent variable can take, except that it must be between 0 and 1. However, the use of OLS to model vote share makes an a priori assumption that the dependent variable *cannot* take on (possible) values of 0 and 1. Therefore, we are of the opinion that it is more prudent to use the fractional logit model to model proportion data, as compared to the OLS.

Pepinsky further questions the fractional response model by stating that "most political scientists use OLS to model vote shares [because] fractional regression methods rarely change substantive conclusions unless vote shares of zero appear frequently in the data," suggesting that the fractional logit model is only useful if a huge number of zero observations is present in the data. Far from it, Papke and Wooldridge (1996, 619–632) highlight that the fractional response model can help avoid the use of ad hoc transformations to handle data at the extreme values of 0 or 1—an added benefit of the model. Hence, the fractional response model was not designed to merely handle the extreme values of 0 or 1, but it has an advantage over earlier methods/models when handling such extreme values. Therefore, there is no restriction for the use of the fractional response model to be only for cases where there are 0s or 1s observed in the dependent variable.

Nevertheless, when judging the fractional logit model against the guiding principle of how wrong the model has to be for it to be not useful, the above explanations that we have put forth show that the use of the fractional logit model is not wrong at all. In fact, Papke and Wooldridge (1996, 619–632) argue that the linear regression model is not a good model specification if the dependent variable is bounded between 0 and 1, primarily because the effect of any particular explanatory variable cannot be constant through the range of the explanatory variable. We also reiterate the point in our article that from the theoretical perspective, the predicted values from an OLS regression are not guaranteed to lie in the unit interval, although we note from Pepinsky's empirical analysis that none of his predicted values exceed the unit interval. However, our earlier explorations and considerations of other model specifications that applied OLS did produce predictions that exceeded the unit interval. Without going into the details of a particular model specification that we considered in early stages of this research, Figures 1 and 2 show the predicted values of the vote share to BN using OLS and fractional logit, respectively. It is clear that the fractional logit model can help constrain predicted values to be between the unit interval, but the OLS cannot. This example has also shown that producing sensible predictions is conditional on the model specification, to which we now turn our attention.

Model Specification and Ethnic Variables

In Pepinsky's article and comment, strong arguments and empirical evidence were put forth to question our model specification and the use of ethnic population total. In particular, Pepinsky puts forth a strong argument to use ethnic proportions, and to use a model specification that considers each of the ethnic variables in turn. Instead of tackling all the issues raised from the beginning to the end, we will demonstrate how from the outset, Pepinsky's model specification has limitations, therefore casting doubt on his subsequent analyses.

Figure 1 Predicted Vote Share to BN via OLS

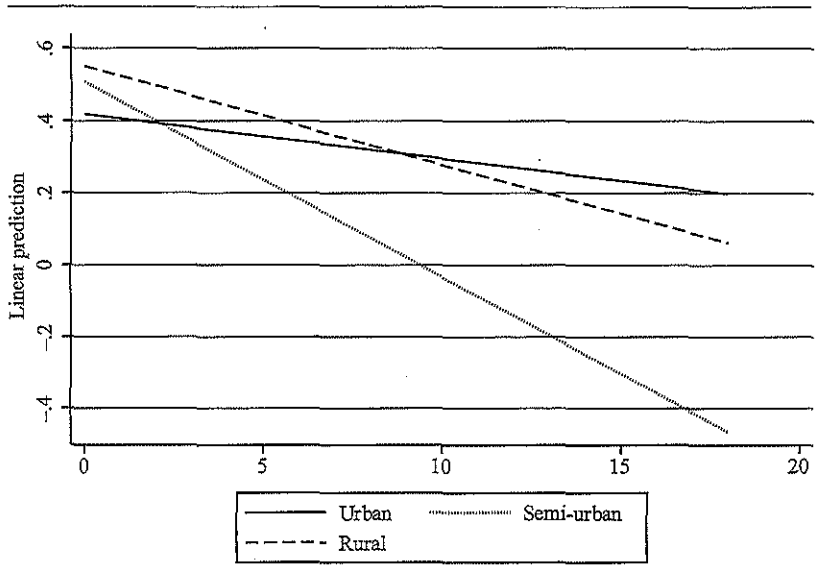
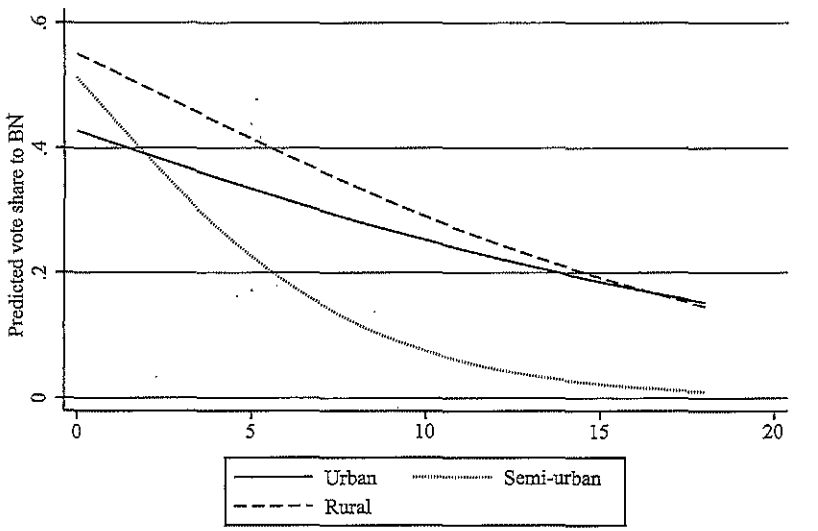


Figure 2 Predicted Vote Share to BN via Fractional Logit Model



Pepinsky has eloquently described the challenges in accommodating for the nature of the ethnic structure in statistical modeling. In short, statistical challenges arise from the fact that this set of compositional data faces a constraint whereby the ethnicity population shares must sum to one, that is, $F_{\text{Bumi}} + F_{\text{Chinese}} + F_{\text{Indian}} + F_{\text{Others}} = 1$. As an alternative to our approach of circumventing this challenge by using ethnic population total, Pepinsky proposes, and subsequently applies, "a simple, theoretically appropriate, and statistically sound modeling strategy for testing the effects of ethnic population shares on BN vote shares," whereby he estimates four separate baseline regressions, with each regression including the ethnic proportion of one ethnic group only. The baseline model he considers is therefore represented as such:

$$BN\ Share = \beta_0 + \beta_1\ \%\ Ethnicity_i + \delta \boldsymbol{D} + \varepsilon$$

where $\boldsymbol{D}$ is a vector of state fixed effects, and ε is an error term. Pepinsky then claims that by doing so, it "preserves the substantive hypothesis about the predictive effects of ethnicity on BN votes, violates no assumptions about coefficient interpretability due to compositional data problems, and can be extended in a straightforward manner to interaction models." Pepinsky's claims are true—only if the sample space is in the real Euclidean space, which in this case, it is not.

The Case Against Using Proportions in Regression Modeling

Ethnic proportions are compositional data that are *constrained*, and the *components* of the composition must sum to a given constant, which in our case would be either 1 or 100 percent. Accordingly, this data structure is radically different from that of unconstrained data; statistical methods designed for unconstrained data are therefore inappropriate for application to (constrained) compositional data.[2] Therefore, the correct way to model compositional data is to *remove the constraints* of the compositional data via a transformation,[3] perform traditional statistical methods (e.g., OLS) on the *transformed vectors*, and then transform the results back into the original space (Wang et al. 2013).

Pepinsky sought to offer a simple modeling alternative to remove the challenges in dealing with compositional data by including only one ethnic proportion variable in his regression model while deliberately excluding the other ethnic proportion variables. However, this approach does not remove the constraint on compositional data at all.[4] In addition, Aitchison (1986), who made huge advances in this area in the 1980s,

warns against adopting this approach.[5] Subsequently, any regression technique that is applied to the original untransformed compositional variable may give rise to misleading inferences (Hron, Filzmoser, and Thompson 2012, 1115–1128).

The Use of Ethnic Population Total

In recognizing the correct sample space of compositional data, our earlier works attempted to perform the isometric log-ratio transformation as per Hron, Filzmoser, and Thompson (2012). However, we decided against it because of the following:

- It would make the article too technical, distracting the reader from the political issues at hand.
- Interpretation of isometric log-ratio transformed variables is difficult, even in linear regression models, thereby making it hard to make useful inferences.
- No work has been done on how the isometric log-ratio transformation can be performed on quadratic variables and for interaction variables.

In lieu of the above, we decided to go with ethnic population totals as our measure of ethnicity, as the sum constraint would at least somewhat be removed. However, we acknowledge that this is not the best way to model ethnicity, which Pepinsky has correctly and strongly pointed out. Nevertheless, in our opinion, it is the better choice to model the data.

In referring to the guiding principle, again, as to how incorrect the model has to be for it to be not useful, we are of the opinion that leaving the original compositional variables as they are, while leaving out some parts from the regression model, is more incorrect than our approach of using ethnic population totals, which at least attempts to remove the sum constraint. Moreover, Pepinsky's inclusion of one ethnic variable in the regression, while leaving the rest of the ethnic variables out of the regression and effectively moving them into the regression error term, raises questions about potential endogeneity and omitted variable biases.

To address the point raised by Pepinsky—that our use of ethnic population totals has the effect of changing the research question at hand from the analysis of the effect of *ethnic composition* to *ethnic population totals*—let us consider the linear population regression function:

$$E(y|\boldsymbol{x}) = \beta_0 + \beta_1 Bumiputera_i + \beta_2 Chinese_i + \beta_3 Indians_i + \beta_4 Others_i$$

where *Bumiputera*, *Chinese*, *Indians*, and *Others* represent the respective number of voters in each of those ethnic groups. If we consider, for example, the coefficient β_1, it is interpreted as the expected change in vote share to BN from an increase in the number of Bumiputera voters, while holding the number of voters in all the other ethnic groups constant. Therefore, the interpretation is still, to some extent, in terms of relatives (or proportions). It is only if the population regression function is presented as

$$E(y|\boldsymbol{x}) = \beta_0 + \beta_1 Ethnicity_i$$

where i = *Bumiputera, Chinese, Indians and Others*, that Pepinsky will then be correct to say that we would have only been examining the effect of ethnic population totals. In this case, β_1 would be interpreted as the expected change in vote share to BN from an increase in the number of Bumiputera voters. The latter case would then have been examining the effect of ethnic population totals instead. This is, therefore, one of the merits of including all four ethnic groups in our model specification.

To conclude the matter on technical specification issues, the estimation of the fractional logit model and the use of ethnic population totals as variables in our model specification are fully justified. The results are therefore credible and are subsequently useful to draw insightful inferences.

Urbanization and Ethnicity in GE13 Outcome

Turning to his other comments, Pepinsky asserts that "even after decades of urbanization, Malay voters still tend to be rural voters." While it is true that rural voters in Peninsular Malaysia tend to be Malay voters, the converse is not necessarily true. With the rapid rural-urban migration and a higher Malay population growth, an increasing proportion of urban voters are Malay voters (Tey 2012). Many urban constituencies, especially in the east coast states, Kuala Lumpur and Selangor, have either Malay majority or plurality. As this trend continues in the next few decades, the Malay/rural versus non-Malay/urban paradigm that has underpinned much discussion on Malaysian politics in the past may need to be reviewed. In fact, a major motivation for our article is to explore the ramifications of this trend on voting patterns.

Table 2 shows the electoral outcome in parliamentary seats that are classified as urban by Politweet, and where Malays make up more than 50 percent of the electorate.

Of the fourteen parliamentary seats shown, ten were won by PR and only four by BN. If we follow Pepinsky's argument to disregard the urbanization effect and assume that a higher proportion of Bumiputera voters entails a higher proportion of vote share for BN, most of the

Table 2 Voting Results from Urban Seats with Bumiputera Majority

Voting District No. and Name	Majority	Total Number of Voters	Bumiputera Voters		Chinese Voters		Indian Voters		Other Voters		Winning Party
			% of Voters	No. of Voters	% of Voters	No. of Voters	% of Voters	No. of Voters	% of Voters	No. of Voters	
P.125 Putrajaya	5,541	15,791	95.5	15,080	1.1	174	3.1	490	0.3	47	BN
P.036 Kuala Terengganu	10,785	94,406	89.2	84,210	10.1	9,535	0.6	566	0.1	94	PR
P.021 Kota Bharu	15,970	81,268	82.9	67,371	15.8	12,840	0.9	731	0.4	325	PR
P.108 Shah Alam	10,939	99,957	70.5	70,470	14.9	14,894	14.2	14,194	0.4	400	PR
P.119 Titiwangsa	866	55,282	68.7	37,979	19.9	11,001	10.4	5,749	1.0	553	BN
P.083 Kuantan	4,515	56,280	62.9	35,400	33.1	18,629	3.8	2,139	0.2	113	PR
P.009 Alor Star	1,873	69,009	61.3	42,303	33.6	23,187	4.7	3,243	0.4	276	PR
P.118 Setiawangsa	1,390	62,309	57.8	36,015	30.4	18,942	11.2	6,979	0.6	374	BN
P.121 Lembah Pantai	1,847	72,396	55.9	40,469	22.8	16,506	20.0	14,479	1.3	941	PR
P.116 Wangsa Maju	5,511	67,775	54.5	36,937	36.2	24,535	8.8	5,964	0.5	339	PR
P.124 Bandar Tun Razak	11,832	90,993	53.5	48,681	37.4	34,031	8.7	7,916	0.4	364	PR
P.137 Bukit Katil	5,447	99,438	53.1	52,802	40.8	40,571	5.7	5,668	0.4	398	PR
P.160 Johor Bahru	10,134	96,321	52.1	50,183	42.6	41,033	5.0	4,816	0.3	289	BN
P.109 Kapar	23,790	144,159	51.8	74,674	33.6	48,437	14.4	20,759	0.2	288	PR

Note: Urban seats as identified by Politweet.

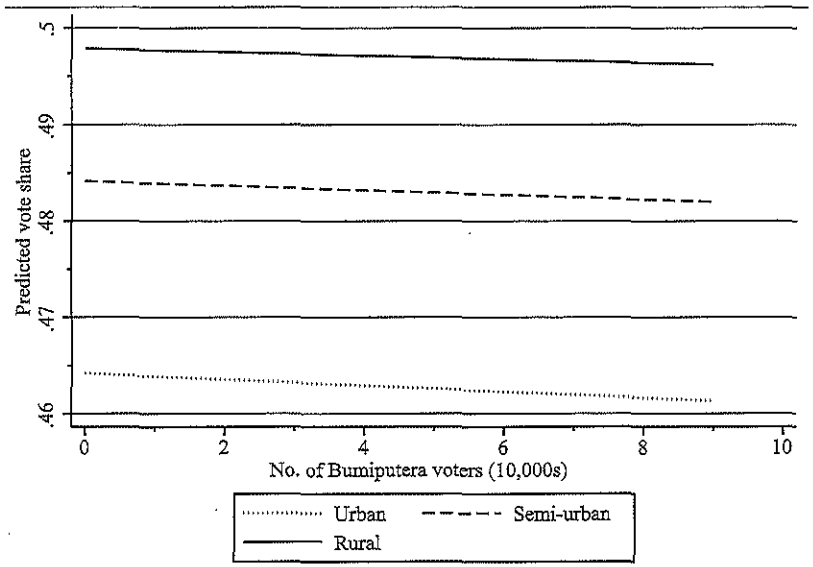
parliamentary seats in Table 2 should have been won by BN. It is also interesting to note that of the four parliamentary seats won by BN, two of them (Titiwangsa and Setiawangsa) were won with majorities of less than 4 percent of the total number of registered voters. In the case of PR, only two (Alor Star and Lembah Pantai) out of the ten parliamentary seats won do not exceed the 4 percent threshold. This analysis suggests that the urbanization effect in determining election outcomes should not be easily discounted. However, as we are using aggregate data in our analysis, there are possible outliers that go against the underlying trend. Specifically, the other two urban parliamentary seats won by BN with large majorities listed in Table 2 are Putrajaya (also highlighted in Pepinsky's commentary) and Johor Bahru. The Putrajaya parliamentary seat encompasses the new administrative capital of Malaysia and Malay voters registered in that constituency are predominantly government civil servants. As for the case of Johor Bahru, BN candidate Datuk Shahrir Samad's personal popularity may have played an important role in ensuring BN's large majority in this urban constituency.

Bumiputera Support for BN

In contrast to Pepinsky's findings whereby a higher proportion of Bumiputera voters corresponds to a higher predicted mean vote share to BN (Pepinsky 2015, Figure 6), our results indicate that Bumiputera support for BN ranges from 46 percent for an urban seat to just below the 50 percent threshold for a rural seat. Bumiputera support remains stable at these percentage levels across increasing numbers of total Bumiputera voters, irrespective of the seat urbanization classification. We reproduce Figure 3 in our article as Figure 3 in this reply.

What is revealing about these results is the fact that even for rural seats, the predicted average BN percentage vote share is less than 50 percent. This finding is surprising, as we had expected rural Bumiputera support for the BN to be significantly higher than 50 percent. There are two possible factors that may have contributed to the lower than expected level of support. First, since we use aggregated data, the support that Parti Islam Se-Malaysia (PAS) receives in its traditional stronghold states of Kelantan, Terengganu, and Kedah may mask a higher level of support that the BN receives in the rest of the rural areas of Peninsular Malaysia. It also can be arguably inferred that such levels of support are due to the presence of out-of-town Bumiputera voters who return to their respective rural constituencies to cast their ballots. These Bumiputera voters work in cities located in urban areas and their mindsets are attuned to issues

Figure 3 Predicted Vote Share for BN in Urban, Semi-Urban, and Rural Constituencies for Varying Numbers of Bumiputera Voters



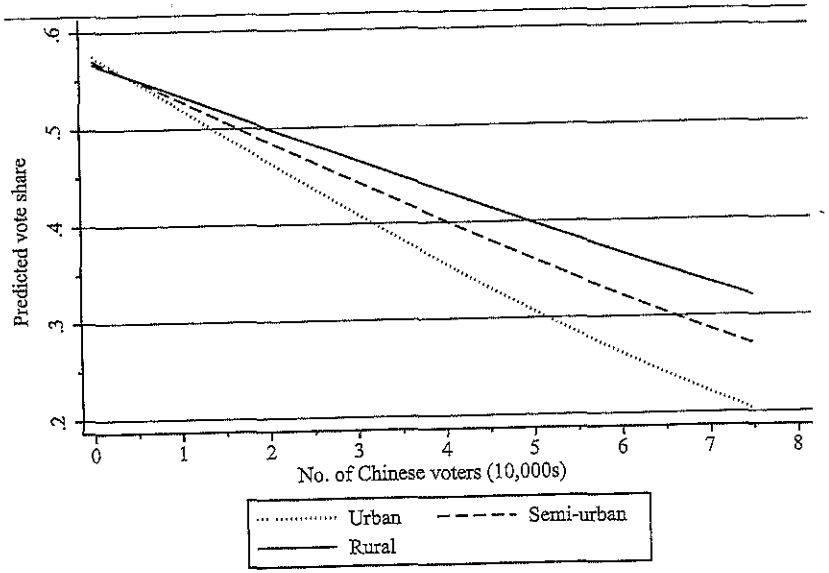
that affect urbanites. They may elicit the mentality of "urban chauvinism" as espoused by Thompson (2013). In conclusion, we agree with Pepinsky's view that BN has firm support of the Bumiputera electorate in the rural belt. However, we find this support to be less than 50 percent, even for rural seats. This means that some level of support has to come from the Chinese voters, which constitutes the second part of our findings. We now elaborate on our findings on rural Chinese voters and also illustrate an added advantage of our econometric model specification.

Chinese Voters Helped BN Cross the Finish Line

In our article, we found that Chinese voters voted overwhelmingly for PR in urban seats. However, the significant results we would like to reiterate are depicted in Figure 4, which is a reproduction of Figure 4 in our article. It also depicts an added advantage of using total ethnic population rather than ethnic population proportions, as put forth by Pepinsky.

Figure 4 clearly indicates that Chinese support for BN declines drastically when the total Chinese voter population rises for all constituency

Figure 4 Predicted Vote Share for BN in Urban, Semi-Urban, and Rural Constituencies for Varying Numbers of Chinese Voters



types, with urban constituencies showing the steepest decline. The surprising result is the level of Chinese voter support for BN when their numbers are small. Support for BN is around 54 to 57 percent when their numbers are around 5,000. This support is crucial for BN because Bumiputera support is slightly less than 50 percent, even in rural areas, as we have shown. These Chinese voters essentially helped BN cross the finish line. If we had used the proportion of ethnic voters, as argued by Pepinsky, we would not have uncovered this new contribution to the Malaysian politics literature.

Conclusion

We have clarified that although our initial intent was to find which variable, ethnicity or urbanization, was the dominant factor in explaining the thirteenth Malaysian general election, our results have shown that both variables are important in determining BN vote share. We do not take a "horserace" perspective, as put forth by Pepinsky, but rather have shown that both variables operate in unison, with the *Chinese-Urbanization* factor being a dominant influence on vote share to BN. We have also ar-

gued that Malay voters need not be predominantly rural. The econometric methodology we have used in our article breaks new ground toward contributing to the vast literature on Malaysian politics. The surprising results on Bumiputera and Chinese voters' support for BN in rural areas shed new insights on voter behavior that could never have been uncovered by OLS methodology. Much more can be done to improve our modeling of Malaysian voters' behavior. Future research should incorporate control state variables, as per Pepinsky's analysis.

Future work at the micro level can be directed at explaining the observed behavior of Chinese voters when they make up a small minority of the electorate in any particular seat. Reliance on the government may be an important contributing factor. However, more research needs to be done to understand the motivations for this behavior.

Our results present a working hypothesis that can be answered only in the definitive if we look at micro-level data: looking at the voting patterns in individual voting streams across voting districts that are categorized as urban, semi-urban, or rural. However, comparing those who actually vote against the electoral roll to determine ethnicity is indeed an impossible task for any social scientist, given that ethnicity is not listed for each voter on the electoral roll. The approach taken in our original article seems to be the better option. Other data on the electoral roll can open up more avenues for research within the context of Malaysian politics. Electoral rolls provide information on the age and gender of the voter. Future research can tap these data to examine the voting patterns of women as well as the voting patterns of various age cohorts (see, for example, Khor 2014, 89–121).

Notes

1. http://robjhyndman.com/hyndsight/statistics-vs-econometrics/.

2. The sample space that compositional data occupy is referred to as the "Aitchison geometry on the simplex" (Hron, Filzmoser, and Thompson 2012, 1116). In contrast, unconstrained data are associated with the real Euclidean sample space. In particular, there is a nonlinear relation between the Euclidean space and the Aitchison geometry, therefore making it inappropriate for standard statistical methods designed for unconstrained data to be applied *directly* to constrained compositional data.

3. Some of the transformations developed over the decades include the additive log-ratio transformation, the centered log-ratio transformation, and the isometric log-ratio transformation.

4. The one ethnic proportion variable that remains in the model is still *not* in the Euclidean space.

5. Aitchison classifies those who opt for the nontreatment of compositional variables as "wishful thinkers." In particular, Aitchison (n.d., 111) says, "No problem exists (Gower 1987) or, at worst, it is some esoteric mathematical statistical curiosity which has not worried our predecessors and so should not worry us. Let us continue to calculate and interpret correlations of raw components. After all if we omit one of the parts the constant-sum constraint no longer applies. Someday, somehow, what we are doing will be shown by someone to have been correct all the time."

References

Aitchison, John. "A Concise Guide to Compositional Data Anlaysis." n.d. www.leg.ufpr.br/lib/exe/fetch.php/pessoais:abtmartins:a_concise_guide_to_compositional_data_analysis.pdf.

———. 1986. *The Statistical Analysis of Compositional Data.* New York: Chapman and Hall.

Box, George Edward Pelham, and Norman R. Draper. 1987. *Empirical Model-Building and Response Surfaces*, Wiley Series in Probability and Statistics. New York: John Wiley and Sons.

Gower, John C. 1987. "Introduction to Ordination Techniqes." In *Developments in Numerical Ecology*, ed. Pierre Legendre and Louis Legendre, NATO ASI Series. Berlin: Springer.

Hron, K., P. Filzmoser, and K. Thompson. 2012. "Linear Regression with Compositional Explanatory Variables." *Journal of Applied Statistics* 39, 5: 1115–1128.

Khor, Yu Leng. 2014. "The Political Tussle over Felda Land Schemes—UMNO Strengthens Its Malay Rural Fortress in 13th General Elections." *Kajian Malaysia* 32, Supp. 2: 89–121.

Ng, Jason Wei Jian, Gary John Rangel, Santha Vaithilingam, Subramaniam S. Pillay. 2015. "2013 Malaysian Elections: Ethnic Politics or Urban Wave?" *Journal of East Asian Studies* (this issue).

Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates." *Journal of Applied Econometrics* 11, 6: 619–632.

Pepinsky, Thomas B. 2015. "Interpreting Ethnicity and Urbanization in Malaysia's 2013 General Election." *Journal of East Asian Studies* (this issue).

Tey, Nai Peng. 2012. "Internal Migration in the Klang Valley of Malaysia: Issues and Implications." *Malaysian Journal of Chinese Studies* 1: 40–59.

Thompson, Eric C. 2013. "Urban Cosmopolitan Chauvinism and the Politics of Rural Identity." In *Cleavage, Connection and Conflict in Rural, Urban and Contemporary Asia*, ed. Tim Bunnell, D. Parthasarathy, and Eric C. Thompson, ARI—Springer Asia Series. Dordrecht, Netherlands: Springer.

Wang, Huiwen, Shangguan Liying, Wu Junjie, and Guan Rong. 2013. "Multiple Linear Regression Modeling for Compositional Data." *Neurocomputing* 122: 490–500.