The Unintended Consequences of Internet Diffusion: Evidence from Malaysia

Luke Miner
Department of Economics, LSE

January 2012

Job Market Paper

Abstract

Can the introduction of the internet undermine incumbent power in a semi-authoritarian regime? I examine this question using evidence from Malaysia, where the incumbent coalition lost its 40-year monopoly on power in 2008. I match IP addresses with physical locations to construct a measure of internet growth in Malaysia from 2004 to 2008. Using an instrumental variable approach to account for endogenous internet placement, I find that areas with higher internet penetration experience higher turnout and higher turnover, with the internet accounting for one-third of the 11% swing against the incumbent party in 2008. In fact, the results suggest that, in the absence of the internet, the opposition would not have achieved its historic upset in the 2008 elections.

JEL Classification: D72, O33, L82, R11

Keywords: internet diffusion, political economy of the media, Malaysian elections.

Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK. Email: l.miner@lse.ac.uk

[1] I am extremely grateful to Robin Burgess, Greg Fischer, and Gerard Padró i Miquel for their encouragement and support throughout this project. Jayanath Appudurai, Oriana Bandiera, Tim Besley, Graham Brown, Premesh Chandran, Bill Cheswick, Edward Clayton, Charif El-Ansari, Maitreesh Ghatak, Norizan Ghazali, Edmund Terrence Gomez, Steve Gibbons, Hayati Harudin, Jason Ketola, Erik Kline, Koay Hock Eng, Valentino Larcinese, Alan Manning, Thomas J. Mather, George Michaelson, Kian Ming Ong, Yeong-Chuan Lim, Henry Overman, Torsten Persson, Steve Pischke, Danny Quah, Ramasamy Ramachandran, Yasuyuki Sawada, William Stevenson, Ibrahim Suffian, Dr. Fadhlullah Suhaimi Abdul Malek, Tan Tze Meng, Hansen Toh, Tom Vest, Georg Weizsäcker, and Dieter Zinnbauer provided valuable advice and comments.

1. INTRODUCTION

The internet has grown enormously over the past two decades: from its DARPA roots in the U.S. Department of Defense, to today where it is a near-ubiquitous method of communication and information exchange in developed countries and—thanks to the rise of mobile telephony—a rapidly expanding technology in developing countries. Since its inception, much has been made of the internet's potential as a democratizing force that frees information from the control of governments, implodes the distance between users around the world, and provides access to new viewpoints. Indeed, the internet's ability to provide unfiltered access to information has caused consternation among many governments. This response has been notable in China, which has invested billions in keeping the internet under tight rein. Social media has also been identified as a driving factor behind protests the world over, such as the recent revolutions across the Middle East. Despite a wealth of anecdotal evidence, however, little quantitative work has been conducted to test the ability of the internet to foster democratization.

Malaysia serves as a particularly compelling test case in this regard. First, the ruling coalition, the Barisan Nasional (BN), enjoyed veto-proof control over all branches of government from 1969 to 2008. Although Malaysia holds regular democratic elections, the BN maintained power through strict controls on the judiciary, the police, and, importantly, the mass media.

Second, the BN's hold on power was so secure that it initiated an aggressive information and communications technology (ICT) led development strategy, based on an uncensored internet. The government has invested heavily in the ICT sector since 1996 as a means to promote growth and enjoys a very high rate of internet penetration—60% as of 2008. At the same time, to attract foreign direct investment (FDI), the government pledged not to censor the internet.

Third, since the internet is uncensored, it has become home to a vibrant opposition blogosphere and a number of popular, independent news sites. In March 2008 the BN lost its two-thirds majority in parliament for the first time since 1969, as well as control of 5 out of 13 states. In the aftermath, commentators argued that the internet played a leading role in this outcome by providing access to alternative viewpoints. In the rush to promote an information economy, the government overlooked the consequences with regard to political control. This paper tests whether internet penetration influenced voting behavior in

Malaysia, focusing on the 2004 and 2008 elections.

I develop a simple model to understand the mechanism by which the internet influences election results. The model, building on Besley and Prat (2006), presents a retrospective voting framework in which the incumbent decides whether to buy off the media. This model shows how the internet can influence electoral outcomes in regimes that use mass media control to ensure reelection. The key insight is that the internet can weaken the ruling party's hold on power by undermining its ability to suppress negative information on candidate type. The model's main prediction is that areas with relatively higher internet connectivity will experience lower vote shares for the incumbent party and higher turnover.

An important contribution of this paper is a novel measure of internet penetration, which can be applied to almost any country. For most countries there are no geographically disaggregated measures of change in internet access across time. To address this problem, I use a dataset that maps all of the IP addresses in Malaysia to approximate geographical locations. I aggregate the data up to the yearly period to deal with changes in assignment location across months. Next I use inverse-distance weighting interpolation to convert the data from the city level to the state legislature district level. Finally, I normalize by the number of eligible voters in a district to create the final measure. I find that this measure performs well when tested against census data from 2004.

To address problems of endogenous internet placement and confounding political trends, I instrument for internet growth. I calculate the shortest distance from each electoral district to the backbones of Malaysia's main Internet Service Providers (ISPs). An increase in distance to the backbone leads to higher costs of supplying internet connectivity (e.g., digging new trenches and laying cabling). This provides exogenous variation in internet supply across districts. I exploit differences across ISPs in terms of geographical constraints on the placement of their backbones to argue that distance to the backbone is unlikely to affect voting outcomes directly, conditional on covariates. The identifying assumption is that conditional on baseline district characteristics (ethnic distribution, GDP per capita, population density) distance to the backbone does not affect change in vote share independently of growth in internet access.

Based on the identifying assumption, I show a large, causal effect of internet growth on election results: that the internet can explain about one-third of the 11% drop in support for the BN from the 2004 to the 2008 election.

I run a number of checks on my identifying assumption. I show that distance to the backbone is uncorrelated with election swings in the pre-internet period. By controlling for distance to roads and road density, I find that these instruments do not proxy for distance to roads. Lastly, I show that the instruments are not capturing the effect of distance to railroads: when I drop the backbone that runs along the railways, the result is unchanged.

Next I examine the effect of internet growth on the turnover of politicians in the incumbent party. I show that internet growth led to increased turnover, if baseline internet access is sufficiently high.

Finally, I test to see whether the internet affected voter turnout. I find that a one standard deviation increase in internet growth corresponds to a 1.5% increase in participation.

To the best of my knowledge, this paper is the first to measure the internet's effects on elections. It relates most directly to a large and growing strain of political economics literature on the complex relationship between media, government, and voters. From an empirical standpoint, Besley and Burgess (2002) provide evidence of a positive effect of the mass media on government responsiveness to natural disasters. Reinikka and Svensson (2011) show how a newspaper-based information campaign on education spending in Uganda improved education outcomes. Snyder and Strömberg (2010) exploit variation between newspaper markets and congressional districts to identify positive effects of newspapers both on voters' knowledge of their representative and on federal spending in their district. In terms of broadcast mediums, Strömberg (2004) finds access to the radio significantly affected public spending during the New Deal. DellaVigna and Kaplan (2007) find that the emergence of the conservative Fox News Channel had a large impact on the 2000 U.S. Presidential elections.

While much has been written about the traditional media, very little empirical work examines the effect of internet media. Gentzkow and Shapiro (2011) look at the effects of the growth of new media on ideological segregation in the U.S.A. I believe this paper is the first to provide causally interpretable evidence of the effect of internet media on election outcomes and the first paper to concentrate on a developing country.

In terms of theory, this paper relates to Mullainathan and Shleifer (2005), which finds that increased competition in the media market could lead to increased bias as newspapers slant their news toward their readerships' priors, and to Baron (2006), which ties increased in bias to journalists' career concerns. This paper draws heavily from Besley and Prat (2006), which presents a theoretical framework for government capture of the media and shows how increased competition in the media market can yield better information on candidate quality and increased turnover. I extend their model by differentiating between traditional and web-based media. In modeling the effect of internet-based media, this paper relates to Edmond (2011), which presents a model of media and regime change that distinguishes between print and broadcast media and online social media.

More broadly, this paper relates to literature on the effects of information technology on development. Jensen (2007) looks at the effects of the introduction of mobile phones on fish markets in Kerala; Goyal (2010) similarly analyzes the effects of internet kiosks on crop prices in Madhya Pradesh; and Jack and Suri (2011) explore the impact of mobile payment on informal risk sharing.

Finally, this paper's empirical strategy pertains to literature exploiting geographic variation for identification. Geography has been used to identify the effects of dams (Duflo and Pande, 2007); electrification (Dinkelman, 2012; Barham, Lipscomb, and Mobarak, 2011); social capital (Olken, 2009; Paluck, 2009); ethnic violence (Yanagizawa-Drott, 2010); and the long-term effects of slavery Nunn (2008); Nunn and Puga (2009).

I start in Section 2 by outlining a general theoretical framework to help understand the main mechanism at play. Section 3, shows how the model pertains to Malaysia, providing background on politics, media, and the internet. In Section 4, I describe my data sources, before outlining my method for constructing a measure of internet penetration in Section 5. Section 6 presents my empirical strategy results. I start by exploring the strong correlation that exists in the data and then move on to the issue of causality. In Section 7, I examine additional outcomes and then conclude in Section 8.

2. THEORETICAL FRAMEWORK

In this section I develop a simple model for understanding how the internet influences voting outcomes when all conventional sources of information are government-controlled. The model is based on Besley and Prat (2006), extended to account for differences between traditional media outlets (e.g., TV, radio, print) and web-based media outlets. The model shows how increased internet access can lead to strengthened government control over traditional media outlets, but less control over the general flow of information. This weaker control over information in turn will lead to diminishing vote share for the incumbent and increased turnover. The model goes on to show how this equilibrium can hold even when more than half of voters have access to (uncensored) internet-based media.

2.1. *Basic Model*

I use a two-period retrospective voting model. In the first period, an incumbent party is exogenously in power and is of two possible types $\theta \in \{g, b\}$ ("good" and "bad" respectively). A good party will deliver a benefit of one to voters; whereas, a bad party delivers zero benefit. The incumbent party's type is realized at the beginning of time and is good with probability γ.

Voters do not observe their payoffs when deciding to reelect the incumbent party and must rely on media reporting for information on type. Voters are distributed in a continuum of districts, which can vary along two dimensions. First, districts can differ in terms of the fraction of the population with internet access ϕ_j in district j. Second, districts can differ in terms of population size. ψ_j represents relative population size and is the fraction of total population in each district j. The incumbent party must win in a majority of districts to retain power.

The media comprises two firms: a mainstream media firm of type M, which we can think of as encompassing print, television, and radio; and a web-based firm of type W, which encompasses all internet-based news sources without offline counterparts.[1] This could range from web-based news sites to blogs and Twitter feeds.

A key distinction between the mainstream firm and the web-based firm is that the mainstream firm can reach all voters; whereas, the web-based firm's potential audience is limited to the fraction of voters with internet access. The mainstream firm may also have an online presence, but I assume they are more vulnerable to government capture due to their offline core (I provide a justification for this assumption in section 3.2).

If the incumbent party is good, neither firm observes a signal on quality. If the incumbent party is bad, firms receive a signal s that $\theta = b$ with probability $q \in [0, 1]$. Firms then can report either bad news or no news. As in Besley and Prat (2006), I assume that only

---

[1] This model can be extended to multiple firms of each type, without modifying the result.

verifiable information can be reported; it is not possible to fabricate bad news. I also assume that both firms receive the same information.

The media receives two types of payoffs: revenue from consumers and revenue from government payoffs. If a media firm reports bad news, the total audience-related revenue available in the market is a. If no news is reported, this revenue is normalized to zero. Viewers prefer informative news. If only the mainstream firm reports bad news it captures the entire market, earning revenue a and all voters are informed. If only the web-based firm reports bad news, its payoff is limited by internet coverage $a\Phi$ and only fraction Φ of voters are informed. If both media firms report bad news, the mainstream news gets all offline consumers and splits the internet-capable audience equally with internet firms. All voters are informed and payoffs for mainstream and web firms are $a(1-\frac{\Phi}{2})$ and $a\frac{\Phi}{2}$ respectively.

The incumbent party receives a payoff r for staying in office and can manipulate the news before the vote in the second period. This is modeled as a bargaining game between media outlets and the incumbent party. The party can offer non-negative transfers t_M (for the mainstream firm) and t_W (for the web firm). If a media firm accepts the transfer, they agree to suppress their signal and report no news. However, their payoffs for accepting are limited by transaction costs $\tau_i \in [1,\infty)$. For any given t_i, a firm receives only $\frac{t_i}{\tau_i}$. The incumbent's payoff is $r - t_M - t_W$ if reelected and $-t_M - t_W$ if not.

As noted in Besley and Prat (2006), transfers can take a number of forms, from all-out bribery to special perks for firms owned by the same company that controls the media outlet. The transaction costs will differ depending on factors such as legal institutions and the ownership type. They speculate that transaction costs should be lowest for state-owned firms and higher for independently owned media. In the context of Malaysia, history shows that capture is possible for the mainstream media, but that the web media is too costly to capture.[2] For simplicity, I model the prohibitive cost of capturing the web media with the assumption that the transaction costs for the web are infinite: $\tau_W = \infty$.[3]

The model's timing is as follows. We begin with an incumbent party in office whose type is realized with probability γ. If the party is of good type, all media outlets observe no signal. If the party is a bad type, all media firms observe the signal $s = b$ with probability q and

---

[2] I provide evidence for this in Section 3.2.

[3] τ_W doesn't need to be infinite, but rather high enough such that it is never profitable for the incumbent to capture the web: $\tau_W > \frac{r}{a\Phi}$.

$s = \emptyset$ otherwise. The incumbent party observes the signal that the media firms receive and chooses transfers t_i. Each media outlet observes its transfer and decides whether to accept. If it accepts, it reports $\tilde{s}_i = \emptyset$ and receives $\frac{t_i}{\tau_i}$. If it rejects, it reports $\tilde{s}_i = b$. Finally, voting is sincere. Each voter observes media reports and updates the posterior probability that the incumbent is good $\hat{\gamma}$. She votes for the incumbent if $\hat{\gamma} > \gamma$ and for the challenger if $\hat{\gamma} < \gamma$.[4]

2.2. *Results when all districts are identical*

I focus on a perfect Bayesian equilibrium, restricted to pure strategies in which voters always vote sincerely for the candidate they prefer. I start with the special case in which all districts are of equal size with identical internet penetration rates. First, note that a bad incumbent party will never choose to capture either media outlet if $\Phi \geq \frac{1}{2}$, since by assumption τ_W is too high to capture the web media and web access is so widespread that a majority of voters will discover the party's type and vote them out.

Proposition 1 *Assuming the internet is too costly to capture ($\tau_W = \infty$) and all districts are of equal size with identical internet penetration rates, equilibrium in the game overall is of two kinds:*

1. *A bad incumbent party will capture the mainstream media if $\Phi < \frac{1}{2}$ and $r \geq \tau_M a \left(1 - \frac{\Phi}{2}\right)$, but will not capture the web media. The party will win in all jurisdictions.*
2. *Otherwise, a bad incumbent party will not capture either outlet and will be discovered with probability q.*

Proof: see appendix.

To see this, note that since the web firm will never accept, the most that the mainstream firm can earn by deviating and not accepting the transfer is $a(1 - \frac{\Phi}{2})$ as outlined above. Thus, the mainstream firm will accept only if $t_M \geq \tau_M a \left(1 - \frac{\Phi}{2}\right)$ and, by extension, the bad incumbent will choose to make a transfer only if $r \geq \tau_M a \left(1 - \frac{\Phi}{2}\right)$. A notable implication is that the cost of capturing the mainstream media is actually decreasing with Φ as long as $\Phi < \frac{1}{2}$, reflecting the fact that the incumbent need capture only a majority of the market rather than the entire market.

---

[4] As in Besley and Prat (2006) sincere voting is assumed for analytical simplicity.

If the mainstream firm accepts, its expected audience-related revenue is zero, instead of $qa\left(1-\frac{\Phi}{2}\right)$ in the case in which neither firm is captured. In contrast, the web-based media's market share is $qa\Phi$ instead of $qa\frac{\Phi}{2}$. The mainstream media loses viewers to the web-based media.

Turning to voters, if the mainstream outlet is captured, viewers without internet access will receive no information on candidate quality and will reelect the incumbent. If the fraction of voters without internet is at least one-half, the incumbent will be reelected with certainty and turnover will be zero. If the mainstream firm is not captured, a bad incumbent will be discovered with probability q and the signal will be reported. Turnover, then, is simply the probability that the incumbent is bad and that the media receives a signal to this effect: $q(1-\gamma)$

This yields the following implications.

Proposition 2 *Assuming internet is costly to capture and districts are identical:*

1. *If* $\Phi < \frac{1}{2}$ *an increase in internet access leads to*
   - (a) *Lower voting shares for the incumbent*
   - (b) *No change in turnover*
   - (c) *Loss of market share by mainstream media outlets*
2. *If* $\Phi \geq \frac{1}{2}$ *an increase in internet access leads to*
   - (a) *Lower voting shares for the incumbent*
   - (b) *Turnover increases from* 0 *to* $q(1-\gamma)$ *across all districts.*

Proof: see appendix.

2.3. *Extension: internet penetration and population size vary across districts*

In this section I relax the assumption that internet penetration ϕ and population size ψ are identical across districts. I show the sufficient distributional assumptions needed for the incumbent party to capture the mainstream media and win a majority of seats even when internet penetration across the country as a whole is greater than 50%.

First, consider the case in which only internet penetration ϕ is allowed to vary and let its distribution be $f(\phi)$. If $f(\phi)$ is rightly skewed around $\phi = \frac{1}{2}$, it is possible to have a greater mass of districts with $\phi < \frac{1}{2}$, but a long right tale leading to $\Phi > \frac{1}{2}$. Under these conditions, government capture could be sustained with $\Phi > \frac{1}{2}$, since a majority of districts have internet penetration rates lower than 50%. The maximum value of $\Phi > \frac{1}{2}$ for which media capture is still optimal, $\overline{\Phi}$, is increasing in the rightward skew of $f(\phi)$. [5]

If ψ (fraction of overall population per district) is allowed to vary as well, an equilibrium can be sustained where $\Phi > \frac{1}{2}$, the mainstream media is captured and there is no skew in $f(\phi)$. Suppose there is no skew and let $\bar{\phi} = \frac{1}{N} \sum \phi_i$ be the average measure of internet penetration across districts. In the case in which population is identical across districts this value is the same as Φ (fraction of the country's population with internet), but once population varies across districts it is possible for a gap to emerge between these two values $\delta = \Phi - \bar{\phi}$. δ can be thought of as the bias on the electoral effect of the internet due to differences in population across districts. The sign and size of this gap depends on the joint distribution of ϕ and ψ. If internet penetration and population size are positively correlated, δ will be positive and increasing in the covariance between ϕ and ψ.

I call the media "capturable" if any of the above conditions is met, such that the incumbent can win the election by paying off the mainstream media. This leads to the following results:

Proposition 3 *Suppose internet penetration, ϕ, and population, ψ, vary by district, and $\Phi \geq \frac{1}{2}$. The equilibrium in the game is of two kinds:*

1. *A bad incumbent party will capture the mainstream media if: internet penetration is rightly skewed around $\phi = \frac{1}{2}$ and/or ϕ and ψ are positively correlated such that $\phi < \frac{1}{2}$ for a majority of districts but $\Phi \geq \frac{1}{2}$. The party will win in all jurisdictions where $\phi < \frac{1}{2}$ and will retain a majority. The incumbent party will be discovered with probability q in districts where $\phi \geq \frac{1}{2}$.*
2. *Otherwise, the incumbent party will not capture either outlet and will be discovered with probability q.*

Proof: see appendix.

In Section 4.3, I show that the distribution of internet in Malaysia is rightward skewed and internet access is higher in districts with higher fractions of the total population. This

---

[5] Negative skew would lead to the opposite result: mainstream media capture could not be sustained even in a situation where nation-wide penetration rates are less than 50%.

suggests the incumbent party can win elections by capturing the media even when internet access is greater than 50% across the country as a whole.

To summarize, the model's main empirical prediction is that an increase in internet access will cause a decrease in the incumbent party's vote share, in the presence of media capture. Intuitively, the internet allows voters to circumvent media controls, and thus enables them to receive negative signals on candidate quality. The incumbent party's vote share will shrink as an increasing fraction of the population gains access to negative signals. I test this prediction in detail in Section 6.

A secondary implication of the model is that internet growth will yield higher turnover in districts where access is above the 50% threshold, but have no effect in districts with low levels of internet access. Section 7 finds evidence corroborating this prediction.

3. BACKGROUND

In this section I relate Malaysia to the theoretical framework above. I start by outlining Malaysia's political regime, move on to describe the state of the country's media sector, and finish with a discussion of its internet.

3.1. *Political regime*

The model above provides a useful framework for thinking about the internet's effect in Malaysia. Classified variously as "partly free"[6], a "flawed democracy"[7], and a "pseudo-democracy"[8], Malaysia's political regime combines democratic and autocratic elements.

Malaysia is a federation of thirteen states with a parliamentary system of governance. Elections are first-past-the-post and occur for both the national parliament and each state legislature. Since independence, Malaysia has been ruled by the same coalition in various guises, the Barisan Nasional (BN, or the Alliance prior to 1969). Though the BN includes parties representing minorities, most notably the Malaysian Chinese Association (MCA) and the Malaysian Indian Congress (MIC), it is effectively run by the United Malays National Organization (UMNO). The UMNO represents Malay and other "native" ethnic groups, known collectively as Bumiputera (meaning sons of the soil).

---

[6] See Freedom House: <www.freedomhouse.org>

[7] See Economist Intelligence Unit: <www.eiu.com>

[8] See Case (2001)

As in the model in Section 2, the incumbent coalition has captured the print and broadcast media. Media ownership is concentrated in a handful of conglomerates that are controlled by the government, constituent members of the BN, and closely connected businessmen. For example, UMNO founded and controls the Utusan Group, which includes the *Utusan Melayu*, the oldest and most widely distributed Malay daily. Also, Media Prima, the largest media conglomerate in Malaysia, owns the largest English daily, two of the largest Malay dailies, four television channels, and three radio stations, and is itself controlled by business proxies of UMNO.[9] Although the traditional media show some variation in their level of bias, in general they tend to under-report on opposition candidates and downplay scandals.[10]

In addition, strict legal restrictions on media outlets prevent the emergence of any mainstream outlet that is overly critical of the government. First, media firms can only operate with a permit and face tightly controlled distribution. Opposition parties are denied permits to publish newspapers, even though constituent members of the BN control multiple media outlets. In the past, publications with critical views of the BN have quickly lost their operating permits and have either shut down or changed ownership.[11] Second, laws such as the Sedition Act, the Control of Imported Publications Act, and the Official Secrets Act allow the government to censor material with impunity.[12] Finally, the Internal Security Act, enacted in 1960 to fight a Communist insurgency, allows detention without trial for up to two years and can be renewed indefinitely.[13]

As in the model, population sizes vary greatly across districts. Apart from media control, the BN has effectively used redistricting as a means to maintain power. This trend, evident across the country at both the state and the parliamentary levels, tends to grossly over-represent rural areas at the expense of cities. Many rural areas act as “vote banks” for the BN, where allegations of vote-buying abound.[14] The period between 1986 and 2008 alone saw three redelineation exercises, with parliamentary seats rising from 177 to 222 and state legislature seats (excluding Sabah and Sarawak) rising from 351 to 455.

---

[9] See Abbott (2011) for a more detailed discussion of the ownership structure of the media.

[10] See Centre for Independent Journalism Malaysia (2008) or <www.malaysiakini.com/news/168567> for specific examples.

[11] See Kua (1990) for examples.

[12] For example, the Home Ministry censored an article in the July 16th, 2011 issue of the Economist on an electoral reform rally.

[13] As of September 2011, an announcement was made to reform these laws. See Section 3.2 for details.

[14] See Pepinsky (2007)

Malaysia has had opposition parties since independence, but only in recent years have they posed a real threat to the BN's hegemony. The rise of a viable opposition can be traced to the late-1990s, when a split between then-Prime Minister Mahathir Mohamad and Deputy Prime Minister Anwar Ibrahim led to Anwar's sacking and subsequent imprisonment under charges of sodomy and corruption. Anwar, once ejected from government, founded an opposition movement called *Reformasi*. After his imprisonment, the movement coalesced into a political party and, following a name change and a merger, is now known as the PKR (People's Justice Party). The other members of the opposition are the Democratic Action Party (DAP), a secular party backed mainly by Malaysian Chinese, and the Pan-Malaysian Islamic Party (PAS), an Islamist party supported largely by Malays in the north of peninsular Malaysia. The PKR, DAP, and PAS contested the 1999 and 2004 elections as part of the Barisan Alternatif (BA), but disbanded the coalition after dismal losses in 2004.

Their fortunes changed dramatically in 2008. The parties wrested control of 5 out of 13 state houses and deprived the BN of its two-thirds majority in parliament. As we shall see below, there is good reason to believe that the internet played a role in this outcome.

3.2. *The internet and politics*

The second essential parallel to the model lies in Malaysia's treatment of the internet. In stark contrast to print, radio, and television, the internet has never experienced significant censorship.

By 2008, Malaysia had a very high rate of internet access. Internet users comprised 56% of the population, compared to 24% in Thailand and 75% in Japan.[15] Malaysia's high internet penetration rate stems from Mahathir's decision, in 1996, to invest heavily in ICT infrastructure as a way to foster a knowledge-based economy. At the forefront of this effort was the creation of the Multimedia Super Corridor (MSC), a high-tech zone south of Kuala Lumpur. This project entailed large-scale investment in constructing a brand-new "high tech" city called Cyberjaya and two new universities. It formed part of the MSC's primary goal: to attract multinational companies through tax breaks and first-class infrastructure. Most important, to make the location more attractive to FDI, the government signed an internet

---

[15] See World Bank Development Indicators: <data.worldbank.org>

"bill of rights", pledging not to censor the internet.[16]

As a result, the internet is the only platform available for alternate view points and has become an important source for independent news and opposition news and views. As early as 1999, members of the pro-Anwar Reformasi movement used blogs and newsgroups to spread their message, and there is some evidence that the internet influenced the 1999 general elections.[17] Although Reformasi sites ebbed in subsequent years, the opposition continued to dominate the web as opposition lawmakers joined citizen bloggers to try to reach a wider audience. The *Harakah Daily*, a PAS-owned news portal, represents the most ambitious online effort by an opposition party; it effectively allows the PAS to distribute a newspaper without obtaining a permit. In addition, a number of independent online news sites, the most famous being Malaysiakini, tend to favor the opposition. Finally, services such as YouTube, and increasingly Twitter, have made it easy to spread the word about political scandals and protest movements.

In terms of the model, which requires that information be verifiable, the internet's most salient feature is its ability to provide information about scandals that previously would have been suppressed. The best example, the V.K. Lingam video uploaded to YouTube in late-2007, showed high-level officials engaged in judicial fixing for the Supreme Court. The video received millions of hits in a matter of days and erupted into one of the defining issues of the 2008 elections. In line with the model's predictions of effects on the media market, evidence suggests that the popularity of the captured media declined. According to a 2010 survey by the Merdeka Center, an independent polling organization, only 40% of Malaysians trust the mainstream media, down 20% from a similar poll conducted two years earlier.[18]

An important assumption in the model is that the internet is too costly to capture. This assumption holds true in Malaysia for a number of reasons. From a purely economic standpoint, censorship scares away the FDI needed for Malaysia's ICT-based growth strategy: as user data migrates from the desktop to servers in the cloud, multinationals are loath to expose themselves to governments that try to limit how they can use this data.[19]

---

[16] See MSC Malaysia Bill of Guarantees at <www.mscmalaysia.my.

[17] See Zinnbauer (2003)

[18] See <www.merdeka.org>

[19] For example, Blackberry maker RIM, has fought against attempts by various governments to access their encrypted network. See <http://online.wsj.com/article/SB10001424052748704017904575409093226146722.html>

The internet is also physically more difficult to regulate than other forms of media. In Malaysia, most opposition content is hosted on international platforms such as Google (Blogger and YouTube), Twitter, and Facebook. As Google's exit from China and all platforms' response to the Arab Spring convey, these platforms tend to dislike engaging in self-censorship. As a result, even if independent news sites were shut down, users could still access content hosted abroad and add to it by posting anonymously. Efforts could be made to censor the internet, as in China's "Great Firewall", but they would be prohibitively expensive and ultimately ineffectual. In terms of expense, such censorship would require substantial investment in not only physical capital but also human capital.[20] Internet censoring would also prove futile; it is always possible for users to leak information,[21] browse anonymously,[22] or bypass firewalls—even ones as sophisticated as China's.[23]

There are signs that the inherent difficulty in controlling the internet is starting to have an effect on government policy. In a speech delivered in August 2011, the current prime minister of Malaysia, Najib Razak, stated: "In today's borderless, interconnected world, censoring newspapers and magazines is increasingly outdated, ineffective and unjustifiable."[24] In September 2011, he went on to announce plans to reform Malaysia's media laws and repeal the Internal Security Act, which allows for detention without trial. It remains unclear if this announcement will translate into meaningful reforms.

3.3. *Internet placement*

Official sources state that the primary motives for building ICT infrastructure were, first, to help Malaysia attain the status of a developed nation[25] and, second, to help promote a Bumiputera business class.[26] In practice, only the first of these goals seems to have played a serious role, with geographical costs being equally important.

The current state of Malaysia's ICT infrastructure originated in efforts to liberalize the

---

[20] China has an army of internet police whose sole job is to peruse forums, blogs, search results, etc. for objectionable content. See <http://www.guardian.co.uk/technology/2005/jun/14/newmedia.china>

[21] Either by posting anonymously to services like YouTube, or to organizations like Wikileaks.

[22] Tor is the best known program for anonymous browsing. See <http://www.torproject.org/>

[23] Ultrasurf, for example, allows users within China to circumvent internet filtering by routing their connection through proxy servers.

[24] See <http://www.economist.com/node/21526885>

[25] First seen as early as 1991 in Mahathir's Vision 2020 development policy.

[26] This is laid out as an explicit aim in the National Telecommunications Plan, 1994.

telecommunications sector in the early-1990s. The government listed the public telecom, Telekom Malaysia (TM), on the local stock exchange (and retained majority ownership); it issued licenses for private telecoms; and it established a new, independent regulator. The liberalization process was poorly planned, with a large number of licenses issued in a very short period to well-connected businessmen. Hungry for profits, but lacking experience in the ICT sector, these private operators went on an uncoordinated infrastructure building spree. Mass bankruptcies ensued in the 1997 Asian financial crisis.[27]

The country emerged from the crisis with a high but uneven level of connectivity: redundant infrastructure in some areas, and a lack of basic telephony services in others. To address infrastructure redundancy, the government encouraged consolidation and infrastructure-sharing. Figure 7 gives an idea of the state of Malaysia's internet infrastructure from 2000 onward, showing the three largest internet backbones currently in operation:[28]

1. *Telekom Malaysia (TM)*: TM, the state-owned incumbent, has the most coverage and capacity. It accounts for just over half of all private internet connections between 2004 and 2008.[29] TM also sells capacity to ISPs that lack extensive physical infrastructure. These ISPs complain, however, that the rates that TM charges are too high for them to compete with TM's own services, especially in areas that are not served by other backbones.[30] Interviews suggest a mixed view of government involvement in TM's placement decisions. On the one hand, the government expects TM to perform the bulk of the heavy lifting to bring infrastructure to remote areas. On the other hand, in the wake of costly bankruptcies after the Asian financial crisis, the government has placed increased emphasis on TM turning a profit. Interviews with planning engineers suggest that demand and geography were the primary factors in infrastructure development since 2000.
2. *Time dotCom (Time)*: Time, a private company, has its own ISP geared toward consumers and businesses. Like TM, it sells excess capacity to ISPs that lack physical infrastructure. As shown in the figure, Time covers less area than TM and overlaps

---

[27] For a complete analysis of the liberalization of Malaysia's telecom sector see Salazar (2007) chapter 7.

[28] Malay's fourth major backbone, Fibrecomm, runs along Malaysia's major power lines. However, this could not be included due to a lack of reliable GIS data.

[29] Budde 2009, Malaysia internet Services.

[30] In fact, the governmental authority in charge of policing the ICT industry has found TM guilty of anti-competitive behavior, but has yet to take any action. See MCMC 2005.

almost completely with TM's network. The red points in Figure 7 are landing stations connected by submarine cabling, which provide network redundancy. There is no evidence of government involvement in Time's placement decisions. If anything, Time went against government wishes by over-investing in redundant infrastructure. As a result, the government had to rescue Time from bankruptcy after the financial crisis (when the current outlines of its present network were already set).

3. *Fiberail*: Fiberail's ownership is split between TM and Malaysia's public railway service. As the name suggests, Fiberail's backbone runs the length of Malaysia's major railways, which were completed in 1931. Given its stake in Fiberail, TM uses Fiberail's network extensively. However, Fiberail positions itself as independent from TM, and sells capacity to ISPs and major corporations. Founded in 1995, Fiberail's initial business activities were restricted geographically to companies with points of presence (access facilities) within a narrow corridor around the railway. In 2006, its license changed such that it could operate throughout the country. In February 2006, Fiberail acquired Petrofibre, a fiber-optic network spanning Malaysia's main gas pipelines.[31] It was impossible to include this additional information on the map, however, as reliable GIS data on pipeline locations are not publicly available.

Annual reports, consultant reports, and interviews concur that cost plays a central role in governing placement, and defining a few key terms will help provide a sense of those costs. ISP backbone refers to the trunk lines, nodes, and routers that form the core of an ISP's network. Linked by bundles of fiber-optic cables, which provide high speed and capacity, backbones are constantly upgraded and occasionally expanded. The backbones form only a part of the network that connects a user to the internet, however. When a user logs on to the internet, for example, the signal must first travel along a length of cable (usually copper), which connects the user's location to a local exchange on the edge of the ISP's network. This first step is often called the local loop or the last mile. The signal then travels along a backhaul connection (normally cabling) until it reaches an access point to the ISP's backbone, called a point of presence. Depending on the size of the ISP, the signal may need to pass through several other ISP networks before reaching the internet. Alternatively, the ISP itself may be directly connected to the internet via, for instance, an intercontinental

---

[31] Budde 2009, Malaysia Telecommunications Infrastructure.

submarine cable.

The costs of delivering the internet to consumers can be divided into several categories. First is the cost of installing the backbone. Geographical and legal factors are the main impediments to backbone placement. In terms of geography, costs include digging trenches so that the fiber-optic cabling can be laid underground. These trenching costs depend on the terrain: it is much more expensive to lay fiber-optic through a jungle than alongside a road. All three backbones therefore follow preexisting routes: roads and highways in the case of TM and Time, and railways in the case of Fiberail. In terms of legal impediments, firms must obtain licenses to run cabling and erect infrastructure. Most land-based trunk cabling runs along federal and not state roads, since it is much less costly and time-consuming to secure a license from the federal government than from state governments. There are substantial differences between state and federal roads. Federal roads tend to be larger than state roads. Whereas the bulk of peninsular Malaysia's federal road system was built by the British before independence in 1957, the state road system continues to grow rapidly.

Once the backbone has been laid, plenty of supplementary costs must be incurred before an ISP can deliver its service to consumers. To serve a new area, an ISP must install a local switch and connect it to the backbone via backhaul cable. This step adds further trenching costs, which increase with distance to the backbone. It also entails the costly and time-consuming process of getting permission from local authorities. Even TM, which owned an extensive telephone network before the advent of the internet, faces these costs. TM had to upgrade much of its copper wire to carry data signals, and dig up and replace its backhaul cable with fiber-optics to provide the extra capacity and speed needed to delivery internet.

4. DATA

4.1. *Political Data*

Malaysia is a federation of ex-British colonies. It is split between peninsular Malaysia, which gained independence in 1957 and houses most of the population, and Sabah and Sarawak, two less developed states on the island of Borneo that joined the federation in 1963. This paper uses election data at the state legislature level for the 1986, 1995, 1999, 2004, and 2008 elections. State elections are held at the same time as elections for the national parliament, with the exceptions of the two states on Borneo, Sabah and Sarawak. Sabah harmonized its state elections with the parliamentary elections only in 2004, and Sarawak continues to hold its state elections on off years. The data includes candidate names, parties, and votes along with turnout, the number of eligible voters in a district, the number of rejected votes, and the district's ethnic composition. I have manually entered each set of electoral boundaries into ArcGIS to account for changes in district size and number since 1986.

Figure 1 shows state and parliamentary electoral district boundaries for the 2004-2008 period. No redistricting occurred in this period. Parliamentary district boundaries perfectly match state legislative district boundaries, with each parliamentary district comprised of two or three state districts.

Table I provides summary statistics covering the 2004-2008 period for state legislature districts in peninsular Malaysia, excluding Kuala Lumpur. The 2008 election is marked by a large drop in vote share for the BN and a modest increase in turnout. The number of eligible voters varies significantly across districts, with a mean of 18,000 and a standard deviation of around 7,000.

4.2. *Demographics and geography*

I have complete geospatial data for Malaysia. Figure 2 illustrates clutter data (which classifies all land as either urban, semi-urban, plantation, jungle, inland water, or open) and elevation data (which allows for the calculation of land-gradients). Figure 3 shows the locations of all major roads, highways, and railways in Malaysia. Finally, Figure 4 represents data from the LandScan service, which estimates population distribution at the one square kilometer resolution through a combination of census data and satellite imagery.[32]

Table I helps make sense of the geo-spatial data. State legislature districts, on average, are 21% urban and 50% farmland (rural), with the remaining 29% classified as jungle. Although jungle covers large swaths of the country, the fairly extensive road network spans more than 80,000 kilometers of roads as of 2007.

I have constructed a dataset of controls using the Population and Housing Census of Malaysia for 1980, 1991, and 2000; Malaysia's Household Basic Amenities and Income Survey for 2004; and geographically disaggregated measures of GDP per capita 2005, generated by

---

[32] See <http://www.ornl.gov/sci/landscan/> for details on the construction of this dataset.

the consultancy Booz & Company. Unless otherwise stated, this data is available at the level of Malaysia's 927 census districts, called *mukim*.

Figure 5 shows mukim boundaries alongside state legislative district boundaries. As can be seen, mukim level data does not match up perfectly with state legislative districts. To address this discrepancy, I use the LandScan population data to assign a weight to each one kilometer cell within each mukim. State electoral district values are generated from the weighted sum of these one square kilometer cells.

Malaysia is a multi-ethnic society. Ethnic Chinese, the wealthiest group in Malaysia, comprise 26% of the population. Brought in by the British as indentured servants to work in the country's rubber and palm plantations, Indians currently comprise roughly 8% of the population. The remainder (65% percent) is largely Malay, except for several ethnic groups on Borneo and a few small tribes.

4.3. *Internet*

I use official internet measures from the Population and Housing Census 2000 and the Household Basic Amenities and Income Survey (HBAIS) for 2004. Both datasets provide the fraction of households with internet subscriptions at the mukim (census district) level. As explained in Section 4.2, I use ArcGIS to aggregate the HBAIC data to the legislative district level, which introduces some measurement error.

In the model in Section 2.3, I presented two sufficient conditions for media capture when average internet access is greater than 50%. The first condition is rightward skew in the distribution of internet connectivity by district. Figure 8 shows an approximation of the PDF for internet subscription per household variable alongside the PDF of a normal distribution. As can be seen the distribution is severely rightward skewed. The second condition is a positive correlation between internet penetration and the fraction of total population. Figure 9 graphs a scatter plot of the log of households with internet subscriptions in 2004 against the fraction of total eligible voters in a district. Showing a strong positive relationship between these two variables, this graph implies that districts with larger populations also have higher internet penetration per capita.[33]

---

[33] These results hold for alternate measures of internet penetration for 2004 and 2008 explained in Section 5.

Since the census data does not cover the 2008 period, I turn to two extra sources of data on internet connectivity. The first, the GeoIP City database, is produced by the geo-location company MaxMind. GeoIP City is a service that matches IP addresses to geographical locations, allowing web services to tailor advertisements based on visitor location and to detect fraud. The GeoIP City database comprises monthly data from 2004 to the present and covers virtually all IP addresses in the world.[34] For each IP address assigned to Malaysia, the GeoIP City database provides the name and location of the nearest large city on a monthly basis. Figure 6 shows the spatial distribution of GeoIP data points for 2008. There are 782 locations that appear in the data for 2004 and 487 for 2008. Although the 2008 data has fewer locations, it has roughly twice as many IP addresses, reflecting the enormous growth in Malaysia's internet penetration in the 2004-2008 period.

My second data source comes from APNIC (Asia-Pacific Network Information Center), the regional internet registry responsible for delegating blocks of IP addresses to national internet registries, ISPs, and large companies in the Asia-Pacific region. As such it has a complete record of all IP blocks allocated to Malaysia along with the recipient of the block (normally an ISP) and the date of allocation.

5. CONSTRUCTING A MEASURE OF INTERNET PENETRATION

Since official statistics do not cover the 2004-2008 period, I construct a novel measure of internet penetration, $IPperVoter$, at the state legislature district level.

I use the GeoIP City database and APNIC dataset, outlined in section 4.3, which together allow me to identify: the initial date of assignment to Malaysia, the ISP managing the IP addresses, and the IP blocks location(s) during the 2004-2008 period. $IPperVoter$ is created by aggregating this data up to the electoral district level and then normalizing by the number of eligible voters.

Challenges

In creating this measure I had to address several sources of measurement error:

---

[34] MaxMind does not cover IPv6 addresses. However, IPv6 adoption was infinitesimal in Malaysia at the time.

1. *Change in IP block location over time:* According to Maxmind, in any given month roughly 60% of IP addresses are correctly resolved to a city that lies within 25 square miles of the actual location.[35] To give a sense of what this means, an IP address is not like a static telephone number: although it can remain stable for long periods, it can also change locations without warning. In many cases the IP address is simply reassigned to another computer in the same general area. However, in some cases it may be reassigned to a completely different part of the country. In the context of Malaysia, this would most likely be a problem for smaller cities. Before a city passes a certain level of connectivity, its IP addresses may be routed either to a regional hub or directly to Kuala Lumpur. As a result, this measure will be biased toward larger cities, especially Kuala Lumpur. Indeed, the monthly data is very noisy with smaller cities disappearing and reappearing from month to month.
2. *Change in IP geo-location accuracy over time:* Not only are the locations assigned to an IP address changing over time, but the accuracy of these IP/location pairings are increasing as well. The data for 2004 is particularly unreliable.
3. *Geographical Measurement error:* The dataset only provides the coordinates for the city center; it does not specify the city's limits. Thus, an IP address corresponding to a computer in a small town outside a city (and in a different electoral district) may be incorrectly attributed to the city, introducing further bias toward large cities. Furthermore, since only point data is available, the boundaries between cities are unclear. This can most easily be seen in Figure 6, which presents GeoIP city center data alongside legislative district boundaries. As can be seen, the point locations often appear on the border between two districts, complicating the task of divvying up IP addresses between adjacent districts.

Approach

To address these challenges I construct multiple measures of internet connectivity, each with different strengths and weakness, which I will test against data from the 2004 HBAIS in the next section.

---

[35] This number is periodically updated and can be found at the following address: <http://www.maxmind.com/app/city_accuracy>.

In response to the problem of IP locations changing over time, I test four methods for assigning IP addresses to cities based on data from a year long period.

1. *IPFix:* I limit the sample to IP addresses that never change location over the twelve month period and divide by twelve. This is the most conservative measure, yielding the advantage that the IP addresses are almost certainly assigned to the correct location. The disadvantage is that the majority of the sample is lost with most remaining IPs assigned to Kuala Lumpur.
2. *IPSum:* I sum up the number of IP addresses assigned to each city over a twelve-month period and divide by twelve. This is the second most conservative measure, making equal use of all of the information in the dataset. However, it likely leads to over-counting Kuala Lumpur. For example, there are many cases in which ten or eleven out of twelve observations occur in a single city, with the remainder going to Kuala Lumpur.
3. *IPMax:* I calculate the number of IP addresses assigned to each city for each month. *IPMax* is the value from the month with the most IP addresses. This measure is meant to account for under-counting of smaller cities. The assumption is that, once an area obtains internet access, it is unlikely to subsequently have access physically dismantled. If a location does not appear in subsequent months, this is due to measurement error rather than a subsequent loss of internet connectivity.
4. *IPAvg:* I, again, calculate the number of IP addresses assigned to each city for each month. I then take the average across months when the city appears in the data. This approach is similar to *IPMax*, but aims to correct for possible over-counting of small cities by *IPMax*.

To address the problem of noisy data for 2004, I create two sets of measures for 2004. The first relies on GeoIP data for 2004. The second uses data from 2005 to infer connectivity at the time of the March 2004 elections. I drop all IP addresses from the 2005 data that were assigned after March 2004 and then calculate the four measures listed above. The assumption is that IP blocks assigned before the March 2004 elections are in roughly the same location in 2005.

To diminish geographical measurement error, I smooth the city point-data into a surface, using inverse distance weighting (IDW) interpolation in ArcGIS. IDW interpolation assigns an IP measure to every point in Malaysia: the value at each interpolated point is a weighted sum of the values in the N known points, where closer points get higher weighting. Figure 6 shows an example of IDW interpolation: darker areas have higher numbers of IP addresses.[36] I then calculate the average IP measure for the entire district. Finally, I normalize by the number of eligible voters in the district to generate two sets of measures, one based on 2004 data and the other based on 2005 data: $IPSumPerVoter$, $IPMaxPerVoter$, $IPAvgPerVoter$, and $IPFixPerVoter$.

Choosing the best measure

Table II shows the correlation between the number of internet subscriptions per household (per the HBAIS 2004) and the IP per voter measures outlined above. Specification (1) includes all state legislative districts for peninsular Malaysia and the Borneo state of Sabah. Kuala Lumpur cannot vote in state legislative elections because it was ejected from the state of Selangor in 1974 and made into a Federal Territory. For the sake of completeness, I include Kuala Lumpur's parliamentary districts. Sarawak, the other state in Borneo, could not be included because it holds its elections for state legislatures on off years.

Turning to the results of Table II we see that the measures based on 2005 vastly outperform their 2004 counterparts regardless of the specification. This leads me to conclude that whatever may have been lost by relying on 2005 data is made up for by greater accuracy in the dataset all around. Moreover, $IPFixPerVoter$, which counts only IP addresses that haven't changed location, performs very badly and is in fact negatively correlated with internet subscriptions per household.

In (2) I drop Kuala Lumpur, leading to an immediate increase in correlation for all 2005 measures apart from $IPFixPerVoter$. I interpret this result as arising from the large bias toward Kuala Lumpur mentioned above. Since Kuala Lumpur is measured with such error and does not participate in state legislative elections, I exclude it from my sample.[37] In specification (3), I limit the sample to the 60 state legislative districts in the Borneo state of Sabah. As of 2004, internet usage in Borneo was sparse relative to the rest of the country. As a result, the GeoIP data contains very few data points for Sabah, which in turn leads to very

---

[36] The process can be altered so that values are not calculated for areas over the sea. Since areas over the sea are not included in any calculations, this does not alter the results.

[37] Unless stated otherwise, dropping Kuala Lumpur has no significant effect on my results.

large bias when performing inverse distance weighting interpolation. Thus, it is unsurprising that the correlations for Sabah alone are low in comparison to (2) (Sabah + Peninsular Malaysia) and (4) (Peninsular Malaysia by itself). Sabah's markedly different ethnic makeup and political structure also make comparison with the mainland problematic. For these two reasons, I exclude Sabah from my sample.

In specification (4), which only counts peninsular Malaysia, $IPSumPerVoter$ 2005 greatly outperforms the other measures. As an additional test, for each measure I generate the corresponding 2008 value and calculate growth from 2004 to 2008. All measures estimate negative growth for a few observations (i.e., the number of IP addresses associated with a legislative district is greater in 2004 than in 2008). In this period, the internet penetration rate for the country as a whole increased from 40% to 60%, and the number of IP addresses more than doubled.[38] Thus I interpret instances of negative growth as measurement error. Again, $IPSumPerVoter$ 2005 performs the best, with only 12 districts with negative growth out of 445; $IPmaxPerVoter$ has 21, and $IPAvgPerVoter$ 50. For the remainder of this paper, I use the natural log of $IPsumPerVoter\ 2005$ for my $IPperVoter$ measure in 2004 and the natural log $IPsumPerVoter$ 2008 for $IPperVoter$ in 2008, and drop the 12 districts in which $IPperVoter$ growth is negative. Unless stated otherwise, results are largely unchanged by including these 12 districts. Of the remaining 433 districts in peninsular Malaysia, I drop 6 regions because of uncontested elections in either 2004 or 2008.[39]

Measurement error

Specification (5), corresponding to the sample used in the paper, indicates a correlation of 0.63 between $IPperVoter$ and the benchmark. Although this result indicates a strong correlation, it still leaves a large, unexplained difference between the two measures.

There are several explanations for why this difference occurs. First, since HBAIS data only counts households with internet subscriptions, it most likely underestimates the percentage of households with access to the internet by omitting people who access the internet at work or in internet cafes. $IPperVoter$ should capture IP addresses tied to work and to internet cafes. However, since $IPperVoter$ contains no metric for intensity of usage, it would not take

---

[38] See <data.worldbank.org> for national penetration statistics; APNIC for IP allocation numbers.

[39] I've run a separate set of regressions including uncontested seats, counting BN share as 1 if the BN wins unopposed and 0 if the opposition wins unopposed. Results are more significant.

into account that hundreds of individuals might use the same IP address on any given day.

For reasons stated above, much of this difference likely arises from measurement error. The majority of this error consists of bias toward major cities, first because of the nature of the GeoIP database, which only counts city centers, and second because of the IDW interpolation technique, which treats the centers of major cities as peaks, with connectivity decreasing as we move outwards.

Fortunately, though the measurement error is large, if anything it will lead to an underestimation of my results below. Indeed, in Appendix A, I show formally that the under-sampling of IP addresses farther from major cities yields OLS estimates that are biased toward zero. In Section 6.4, I use instrumental variables that are uncorrelated with the measurement error to derive consistent estimates.

6. EMPIRICAL ANALYSIS

6.1. *Basic Correlations: OLS Estimates*

I start by examining the basic relationship between internet penetration and the BN's share of the vote at the state legislature district level. Figure 10 plots change in voting share for the BN and growth in $IPperVoter$ during the 2004 to 2008 period. As can be seen, there is a strong negative relationship in the raw data, implying that areas with more internet growth are associated with greater negative swings against the BN.

I explore this relationship in more detail by controlling for other characteristics that might affect changes in BN voter share. Let y_{ist} be BN's vote share for legislative district i in state s at time t and $\triangle IPperVoter_{ist}$ be growth in IP addresses per voter:

$$y_{ist} = \alpha_0 + \alpha_1 t + \alpha_2 IPperVoter_{ist} + \rho_i + \delta_i t + \mu_s + \lambda_s t + \varepsilon_{ist} \tag{1}$$

Where ρ_i is the district fixed effect, δ_i is the district trend,μ_s is the state fixed effect, λ_s is the state trend, and ε_{ist} is an idiosyncratic error term. This equation, in turn, can be rewritten in first differences, eliminating ρ_i and μ_s:

$$(2) \qquad \triangle y_{ist} = (y_{ist+1} - y_{ist}) = \alpha_1 + \alpha_2 \triangle IPperVoter_{ist} + \lambda_s + (\delta_i + \triangle\varepsilon_{ist})$$

With two periods of data, it is not possible to estimate the legislative district specific trend δ_i. OLS estimation of equation (2) will be biased as long as $\delta_i + \triangle\varepsilon_{ist}$ is correlated with $\triangle IPperVoter_{ist}$, which we would expect if internet is allocated to areas that are trending for or against the BN for unobservable reasons.

As a first pass, I augment equation (2) with a vector of legislative district covariates (X_{is}) to control for some factors that might affect δ_i:

$$(3) \qquad \triangle y_{ist} = \alpha_1 + \alpha_2 \triangle IPperVoter_{ist} + X_{is}\beta + \lambda_s + (\delta_i + \triangle\varepsilon_{ist})$$

OLS estimates for equation (3) for the 2004-2008 period appear in Table III. The first column, reporting estimates of equation (3) with fixed effects and state trends, indicates a strong negative association between change in BN share and $IPperVoter$ growth. As argued above, ethnicity is a central driver in Malaysian politics, with non-Malays more likely to switch allegiances from 2004 to 2008. Since the Chinese population is wealthier and more urban, it could be that $IPperVoter$ is simply picking up this trend. In column (2) I control for this possibility by adding ethnicity, and although the magnitude of the effect diminishes, it remains strongly significant. In line with anecdotal evidence, Indians swung heavily against the BN relative to Malays.[40]

Another concern is that internet access simply proxies for wealth; the opposition party PKR, for example, derives much of its support from wealthier Malays. In column (3) I add a measure of GDP per capita as of 2005, and again the magnitude drops, but the relationship remains very significant.

Finally, it could be that $IPperVoter$ is capturing the effect of urbanization. As mentioned above, rural districts traditionally support the BN. I control for urbanization of a district

---

[40] The coefficient on percent Chinese is also negative and significant, when percent Malay is the omitted variable.

with the variables population density and the natural log of eligible voters in 2004. Turning to the results of specification (4), the estimated relationship between change in BN share and growth in IP addresses per eligible voter remains unchanged. Meanwhile, there is no evidence of any relationship between population density and voting trends after controlling for district fixed effects and state trends. To give a sense of magnitudes, specification (4) implies that a one standard deviation increase in $IPperVoter$ growth translates to a 1% swing against the BN.

As a further check, I run (3) for the 1995-1999 period, when internet connectivity grew from zero to 15%.[41] As mentioned in Section 3.2, the internet was seen to play a decisive role as early as the December 1999 elections. Significantly, demographic composition of the electoral swing differed in the 1999 election. In 2008, Chinese and Indian voters abandoned the BN in favor of the opposition; whereas, in 1999 minority voters stayed with the BN and the Malay electorate instead split. I create a measure of internet growth from 1995 to 1999, $InternetHH$, which is the natural log of the percentage of households with an internet subscription in 2000 (internet penetration was zero in 1995).

Table IV provides the results. Column (1) shows that, in the absence of controls, a significant positive relationship exists between internet growth and change in vote share between 1995 and 1999. I interpret this as picking up the fact that the bulk of the swing occurred among Malay voters rather than the relatively more connected Chinese. Indeed, adding ethnicity controls in specification (2) renders the relationship strongly negative and significant. In specifications (3) and (4), I control for GDP per capita and population density, and the magnitude of the effect increases.

The results from the 1995-1999 period reinforce the initial finding of a negative relationship between internet growth and BN share of the vote. That this result holds for a completely different measure of internet growth suggests that the result is not merely artifact of the $IPperVoter$ measure. Moreover, since the relationship holds in the presence of a different demographic shift in the electorate, there is less reason to believe that unobserved state trends are driving the result. Notably, that the relationship is larger in magnitude: a one standard deviation increase in percentage of households with an internet subscription implies

---

[41] I cannot run a regression for the 1995-2008 period because of redistricting between 1999 and 2004 and because my measures of internet penetration are different.

a 2% swing against the BN in 1999 (the total swing against the BN in 1999 was 11%). This is most likely arises because $InternetHH$ is measured with less error than $IPperVoter$.

6.2. *Identification Strategy*

Although the OLS estimates demonstrate a negative relationship between internet growth and change in BN share, it remains unclear if the relationship is causal. OLS estimation of equation (3) will not identify the causal effects of internet growth if $\delta_i + \triangle\varepsilon_{ist}$ is correlated with $\triangle IPperVoter_{ist}$. If internet connectivity is allocated more heavily to districts that are trending toward the BN for unobservable reasons (e.g., patronage) then $\hat{\alpha}_{2,OLS}$ would be biased upward toward zero. If anything, however, this would lead me to underestimate the negative relationship. A greater concern is that internet connectivity was allocated to areas that trended against the BN for unobserved reasons, leading to a negative bias in my results.

To deal with these challenges, I use the distances from the centroid of a state to Malaysia's three largest ISP backbones as instruments (Z_{ij}) that are correlated with growth in internet penetration, but uncorrelated with district level characteristics that influence voting behavior. As argued in Section 3.3, cost, which is a major determinant of internet placement, increases in the distance to the backbone. Since the backbones were being built in the 1995-1999 period, the instruments apply only to the 2004-2008 elections. This produces the following system of equations:

$$\begin{aligned}
(4) \qquad \triangle y_{ist} &= \alpha_1 + \alpha_2 \triangle IPperVoter_{ist} + X_{is}\beta_2 + \lambda_s + (\delta_i + \triangle\varepsilon_{ist}) \\
(5) \qquad \triangle IPperVoter_{ist} &= \pi_0 + Z_{ij}\pi_1 + X_{is}\pi_2 + \gamma_s + \tau_{ist}
\end{aligned}$$

The identification assumption is that, conditional on the baseline district characteristics—ethnic distribution, GDP per capita, population density—distance to the backbone does not affect change in vote share independently of growth in internet access. So long as the instruments are also uncorrelated with the bias in $IPperVoter$ toward large cities, they will produce consistent estimates even though $IPperVoter$ is measured with error.

The first endogeneity concern is that the backbones for Malaysia's ISPs run through areas more likely to swing against the BN for reasons that the controls do not capture. Since the backbones pass through Malaysia's most populous regions and cities, the instruments could simply be picking up the direct effect of urbanization on voting trends. I supplement my controls for population density (log of eligible voters, log of total area) with variables based on satellite data. With clutter data on land usage, I create additional controls for the percent of the district that is urban vs. rural vs. jungle. Last, following Burchfield, Overman, Puga, and Turner (2006), I control for the effect of physical topography on urbanization, constructing a variable for the standard deviation of the land gradient.[42]

Another concern with my instrumental variables is that they are picking up the direct effect of Malaysia's major roads and railways on district trends (e.g., via increased trade and exposure to outside information). For now, I include a control for road density but will return to this issue in more detail in Section 6.5.

In sum, I am exploiting exogenous variation in internet supply due to geographical constraints in backbone placement. I include state fixed effects due to differences across states in terms of internet connectivity and sociopolitical factors. Thus I am exploiting within state variation.[43]

6.3. *First Stage*

Table V shows the first-stage estimates for internet penetration growth in state legislative districts, using growth in IP addresses per eligible voter as a proxy for growth in internet access. Column (1) shows estimates of equation (5) with minimal controls for ethnicity. The coefficient on distance to Time, which is highly significant and in the expected direction, suggests that growth in internet access is decreases with distance from Time's backbone.

For Fiberail, both a linear and square term are included. This is meant to capture a non-linear relationship with $IPperVoter$ growth due to restrictions on Fiberail's geographic area of operation until 2006, as mentioned in Section 3.3.[44] The negative coefficient on the linear term can be interpreted in the same way as the coefficient for distance to Time: internet growth decreases as distance increases. The positive square term captures the geographical limitation effect: the relationship between internet growth and distance to Fiberail becomes

---

[42] All regressions have also been run with average land gradient, ruggedness, and the standard deviation of ruggedness with no significant differences. See appendix for details of variable construction.

[43] There are 11 states in the sample and 38.8 legislative districts per state.

[44] The linear term by itself is insignificant. There is no evidence of a non-linear relationship for any of the other instruments

less negative until it reaches a zero threshold.

Distance to TM's backbone remains insignificant regardless of the specification. In fact, even if I run the same set of regressions only including IP addresses assigned to TM, the results are largely the same. This result is consistent with the idea that the government exerted more influence over TM than its competitors, compelling it to build out infrastructure in areas with low demand.

In specification (2), I control for GDP per capita. The size of the coefficients for the instruments decreases yet remains highly significant. Column (3) shows that including controls for population size and density does not noticeably alter the result. In specification (4), I add geo-spatial controls for urbanization. The coefficients of interest decrease slightly in magnitude but maintain their significance. Internet growth demonstrates a negative relationship with population density, but a positive association with percentage of the district that is urban. The most likely reason for this result is catch-up: ISPs had already brought internet service to the most densely populated areas by 2004 and thus had the most room to grow in regions that were urban but more sparsely populated.

Column (5), which includes a control for road density, is my preferred, baseline specification. The coefficients of interest are unaffected, helping to mitigate the worry that the instruments are simply picking up the direct effect of roads on elections. To give some interpretation of the magnitudes here, for every 10 kilometer increase in distance to Time's backbone, $IPperVoter$ growth decreases by 0.18 of a standard deviation.

Finally, in specification (6), I control for BN share in 2004. As shown, I find no evidence of political interference on internet roll-out. Several other results suggest that demand, rather than patronage, was the primary determinant of internet growth. First, there is a strong positive relationship between GDP and internet growth regardless of the specification. Second, there is no significant correlation between ethnicity and internet growth, belying the government's stated goal of ICT investment as a way to promote a Bumiputera middle-class.

6.4. *Instrumental Variable Results*

The IV estimates appear in Table VI. The specifications for (1)-(5) match their first-stage counterparts from Table V. The coefficient on $IPperVoter$ is negative, significant, and of roughly the same magnitude throughout. The Hansen test does not reject the null hypothesis that the instruments are uncorrelated with the error term, lending credence to the identification assumption. The strong and stable coefficients on ethnicity confirm the importance of race in the 2004-2008 elections.

The effect's magnitude drops in column (2), suggesting that $IPperVoter$ in (1) was picking up some of the effect of GDP per capita. The result remains large and significant, however, and in (4) the coefficient of interest returns to its previous size once controls for urbanization are also included.

Column (5), the baseline estimate, includes a control for road density. As shown, the coefficient on $IPperVoter$ stays unchanged. GDP per capita loses its significance altogether, suggesting that it was proxying for urbanization and road density.

To get a sense of the change in magnitudes, for specification (5) a standard deviation increase in internet growth translates to a 3.6% swing against the BN. Putting this shift into context, IP addresses per voter doubled in the 2004-2008 period, while share of the vote for the BN dropped from 63.9% to 52.2%. This implies that internet growth accounted for about a third of the vote swing.

The magnitude is substantially larger than the OLS estimate, which as I show in the appendix, is likely due to measurement error biasing the OLS estimates toward zero.

6.5. *Validity of the Exclusion Restriction*

As a reminder, my identification assumption is that, conditional on baseline district characteristics (ethnic distribution, GDP per capita, population density, road density, percent urban vs. rural vs. jungle), distance to the backbone does not affect change in vote share independently of growth in internet access. Though it is impossible to test this assumption directly, I perform some additional checks to assess its plausibility.

Pre-Internet Trend Tests

The most basic concern is that unobservable characteristics of areas close to the backbone make those areas more prone to swing against the incumbent party in general. I check for this possibility by examining the reduced form relationship between distance to the backbone and swings in previous elections.

Since Malaysia regularly redraws electoral district boundaries, it is not possible to run this exercise for the complete set of preceding elections. Fortunately, the 1969-2008 period has only two other elections—1986-1990 and 1995-1999—in which there was a sizable swing against the BN, and on both occasions boundaries were fixed. I control for ethnicity and population density using the 1991 and 2000 censuses. I also include controls for population density, road density, and land usage based on 2008 estimates. A large expansion in state roads occurred during this time, which introduces error into my road density control. Finally, I control for GDP per capita using a 2005 estimate for 2004-2008, 1996-1999, and 1986-1990. Results do not change if the controls measured with error are dropped.

Table VII shows the results of reduced-form regressions for the 1986-1990, 1995-1999, and 2004-2008 periods. Columns (1) and (2) show a negative and insignificant relationship between vote swing and distance to either backbone in the 1986-1990 period. Turning to column (3), the relationship to distance to Time shifts to positive and significant at the 10% level during the 1995-1999 period. This makes sense since both backbones were partway built during this period. However, as can be seen in column (4), distance to Fiberail remains insignificant. In terms of the 2004-2008 period, column (5) indicates a positive relationship with distance to Time that is significant at the 5% level, and column (6) shows that the linear and square distance to Fiberail variables are jointly significant at the 1% level.

Controlling for alternate channels

A second major concern is that there are unobservable characteristics of areas close to the backbone that only switched on in the 2004-2008 period.

Since Time and TM run along Malaysia's major roads, the greatest cause for concern is that some characteristic particular to the distance to the roads (or an omitted variable driving it) affected voting trends through some channel, which switched on only after the 2004 elections. I believe this possibility is unlikely for several reasons.

First, the backbones travel along Malaysia's federal roads, most of which were built before 1980. Thus, the effects would have had to remain dormant for more than twenty years.

Second, since Time and TM only travel along a subset of federal roads, we can control both for distance to federal roads and distance to major roads. Table VIII presents the results of equation 5 with additional controls for distance to major roads and distance to federal roads. In specifications (2) and (3), I control for distance to major roads and distance to federal roads using distance to Time, distance to Fiberail, distance to Fiberail squared, and distance to TM as IVs. As illustrated, the coefficient on $IPperVoter$ decreases only slightly and maintains its significance regardless of the control. In columns (8) and (9), I run the same set of regressions but use only my road-based IV, distance to Time. In this case, the magnitude stays largely the same. The standard errors increase substantially, but the relationship remains significant at the 5% level.

Finally, I restrict my set of instruments to those based solely on the railway network. In specifications (4), (5), and (6) we see that the coefficient on $IPperVoter$ remains highly significant regardless of the control. Comfortingly, the magnitude stays largely the same as in the case using only road-based instruments.

Since Fiberail travels the length of Malaysia's railroads, it is impossible to include equivalent controls for distance to the railway. However, it is worth noting that the railroad network was completed as early as 1931. Thus, to invalidate the instrument, the effect of proximity to railroads would have to have remained dormant for 75 years and then activate just in time to influence the 2008 elections.

Other issues

Another concern is the possibility of heterogeneous effects of internet access on voting. If the effect of internet access on voting is more highly negative for areas closer to the backbone, my identification strategy would lead to an overestimation of the effect. An example of this scenario is if areas closer to the Time backbone are better able to exploit internet technology through better education. Were that the case, however, we would expect to see a markedly different coefficient on $IPperVoter$ when only distance to Fiberail is used as an instrument. This is because many of the districts near to Time's backbone are far from Fiberail's backbone. However, as Table VIII shows, the coefficient on $IPperVoter$ is largely the same across specifications regardless of the combination of instruments used.

I have run regressions controlling for change in ethnic distribution, eligible voters, and population density between 2004 and 2008. The results are unchanged suggesting that migration is not driving the effect.

6.6. *Additional robustness checks*

Placebo regressions

As an additional check, I test whether internet growth between 2004 and 2008 is higher in areas that were already predisposed to swing against the BN for unobservable reasons. Running OLS on equation (3), I use change in BN share for earlier elections as the dependent variable, but keep $IPperVoter$ 2004-2008 as the independent variable and use the same set of controls. As explained in Section 6.5, this analysis is possible only for two previous elections, 1986-1990 and 1995-1999, with the same limitations to the controls.

The 1986-1990 period is a good test case of whether places that experienced more internet growth between 2004 and 2008 were already more predisposed to swing against the BN. The year 1990 saw an abortive move toward a multi-party system in Malaysia with the BN suffering its worst setback since 1969.[45] It won only 53% of the vote, but managed to retain its two-thirds majority in parliament thanks to gerrymandering.

Panel A of Table IX shows the results of regressing change in BN share from 1986 to 1990 on internet growth from 2004 to 2008. As indicated, the coefficient on $IPperVoter$ proves small and insignificant regardless of the specification. This suggests no correlation between support for the BN in the 1986-1990 period and internet growth in the 2004-2008 period.

Next, I run the equivalent regression for the 1995-1999 period, regressing BN share 1995-1999 on $IPperVoter$ 2004-2008. Recall from Section 6.1 that a robust negative relationship exists between growth in internet connectivity (as measured by the 2000 census) and BN share. Panel B shows that this result does not hold if 2004-2008 measures are used instead. No sign of a relationship between the 1995-1999 election swing and 2004-2008 internet growth appears, regardless of controls. This suggests that the areas with the greatest swing in 1995-1999 differ from areas experiencing the greatest relative growth in internet access in 2004-2008.

---

[45] Two years earlier, divisions in the UMNO, the dominant Malay party within the BN, caused a formal split in the party with a large number of UMNO politicians leaving to form the opposition Malay party *Semangat 46*.

7. ADDITIONAL RESULTS

In this section, I consider the effect of internet diffusion on additional electoral outcomes. I start by checking the secondary prediction of the model: that internet growth leads to higher turnover once internet penetration reaches a high enough level. Next, I check if the internet promoted greater turnout. Last, I predict the outcome of the election if had there been no internet growth over the 2004-2008 period.

7.1. *Turnover*

A secondary prediction of the model is that higher internet penetration will yield higher turnover in incumbent party seats once internet access is sufficiently high. I test this prediction by comparing turnover in seats defended by the BN during the 2008 elections when internet penetration was greater than 50% to turnover in the 1999 elections when internet penetration was below 20%. In contrast to previous specifications the analysis is at the cross-sectional level. I run probit regressions of a BN victory dummy on the level of internet penetration while limiting the sample to districts won by the BN in the previous election. Table X reports the results.

First, I examine the internet's effect on turnover in the 1999 election. Since internet penetration across the country as a whole had reached only 15%, the model would not predict a significant effect on turnover of BN candidates. Turning to the data, I find no turnover in BN-defended seats in the states of Johor and Negeri Sembilan. Since my empirical strategy exploits within state variation, I drop the 68 observations corresponding to these two states. I also drop 7 observations corresponding to the state of Kelantan, where all BN-defended seats fell to the opposition. Specification (1) reports the result of a probit regression for 1999. The effect, positive and insignificant, provides no evidence that low levels of internet penetration substantially affect voter turnover.

In 2008 internet penetration for Malaysia as a whole had surpassed 50%, high enough for the model to imply an increase in turnover. Column (2) affirms this prediction, implying that the BN had less chance of retaining a seat in districts with higher internet penetration. This specification includes the full set of baseline controls plus distance to federal roads. Logit and linear probability specifications yield commensurate results.

To address endogeneity concerns, in columns (3)-(6), I instrument for $IPperVoter$ 2008 using distance to the backbone. The coefficients on $IPperVoter$ 2008 are relatively stable across specifications (3)-(6), but much larger in magnitude than the simple probit case, pointing again to measurement error biasing the result to zero. In column (3), I include all instruments and the effect proves significant at the 10% level. Turning to specification (4), I drop my weakest instrument, distance to TM, and the significance jumps to the 5% level. In column (5), I restrict the instruments to distance to Time and, in column (6), I use only distance to Fiberail and distance to Fiberail squared. Although the point estimates remain similar, the standard errors are much higher, leading to insignificant results. The most likely explanation for the lower significance is that distance to the backbone variables are weak instruments for the level of internet access as opposed to change in internet access. The coefficients on the instruments in the first stage are largely insignificant. Additionally, I run an IV regression for the equivalent linear probability model and include the F-statistics from the first stage. As can be seen, the F-statistics are much smaller than in the case when distance to backbone instruments for internet growth.

7.2. *Turnout*

Although turnout is not modeled, there are both theoretical and empirical reasons to believe that access to better information on politician quality yields increased turnout (e.g., Banerjee, Kumar, Pande, and Su (2010)). I look at the effect of internet diffusion on turnout, focusing on the 2004-2008 and the 1995-1999 periods.

Turnout measures in Malaysia are noisy due to electoral irregularities. Allegations of electoral manipulation range from phantom voters (in which deceased individuals still manage to cast ballots) to multiple votes by the same individual to vote-buying.[46] To address this challenge, I include an extra set of regressions that drop districts with serious irregularities.[47] For the 2004-2008 elections, 13 out of 427 districts are dropped. However, a lack of information on specific examples of irregularities in earlier elections makes it impossible for me to do the same for the 1995-1999 period.

Table XI presents the results. In specification (1), I run equation (3) using change in

---

[46] See Pepinsky (2007) and Hai (2002) for details.

[47] See data appendix for details of irregularities.

turnout as the dependent variable. The relationship between internet growth and turnout is positive but significant only at the 10% level. In column (2), I drop 13 districts with indications of serious irregularities. As indicated, the magnitude of the relationship rises and the significance increases to the 5% level.

Next, I employ an IV strategy, but the instruments prove much weaker in this case; only distance to Time yields significant results. In columns (3) and (4), I use distance to Time as an IV and include the standard set of controls, plus distance to federal roads. In both cases, the size of the effect increases greatly. However, the relationship is significant only if I drop districts with voting irregularities. To give a sense of the magnitudes, column (2) implies that a one standard deviation increase in internet growth leads to a 0.5% increase in turnout. Column (4) implies that a one standard deviation increase in internet corresponds to a 1.5% increase in turnout, or about half the change in turnout between 2004 and 2008.

Specification (5) shows results for identical OLS regression run for the 1995-1999 elections. Magnitudes are similar to OLS estimates (1) and (2), but standard errors are also much greater, leading to lower significance.

7.3. *Predicted results in absence of internet*

To put the previous results in perspective, I predict the outcome of the 2008 election had there not been any internet growth in the 2004-2008 period. Table XII reports the results. Specifications (1) and (2) give the actual result for the 2004 and 2008 elections, respectively. As shown, the opposition captured four additional statehouses in 2008. Specification (3) employs OLS equation (3) to predict results assuming zero growth in $IPperVoter$ from 2004 to 2008. The predicted percent of seats captured by the BN increases in all states, and the BN retains one of the four statehouses lost. Column (4) reports the estimated outcome with no internet growth using the IV specification. In this case, the effect is more pronounced: the BN retains control of three of the four statehouses that switched to the opposition. These results suggest that, without internet growth between 2004 and 2008, the BN's 2008 election setback would have proven fairly modest, amounting to the loss of only one statehouse.

8. CONCLUSION

This paper contributes to our understanding of the effect of internet diffusion on democratization. Focusing on the context in which the traditional media is government-controlled, I have argued that the internet can facilitate evolution toward a two-party system by preventing any single agent from monopolizing information. Malaysia provides a key opportunity to test this idea: ambitious investment in an internet free of censorship coincided with strict controls on all other forms of media.

This paper's central contribution is to quantify the effects of the internet on democratic change, in the context of the huge growth in internet penetration that has accompanied Malaysia's recent electoral upheavals. I present a model, based on Besley and Prat (2006), in which an increase in internet access undermines an incumbent party's ability to guarantee reelection through media control. In line with the model's main prediction, I find that internet growth accounts for one-third of the 11% swing against the BN in the 2008 state elections.

To put this number in perspective, I predict the outcome of the 2008 elections had there not been any internet growth during the 2004-2008 period. IV estimates imply the BN would have retained control of three of the four statehouses that switched to the opposition. Thus the BN's ICT-based development strategy had the unintended consequence of weakening its control.

I go on to test a secondary prediction of the model. I show that internet growth can yield increased turnover if internet access is sufficiently high. Finally, I find evidence that the internet can help spur higher turnout.

Another important contribution is a novel measure of internet growth from 2004 to the present. Such a metric is lacking for most countries in the world, including the U.S.A. My measure of internet connectivity uses IP geo-location data in conjunction with regional internet registry records. I smooth the IP address point data into a surface using inverse distance weighting interpolation and then normalize by population. Finally I check the accuracy against an independent measure of internet diffusion from household census data. This measure is central to the paper's results as it allows me to track internet growth at the state legislature district level. This measure can also extend to research well outside the ambit of this paper. Equivalent IP geo-location data exists for almost every country in the world and is only becoming more accurate as the technology matures.

This paper presents some of the first evidence of the internet's quantitative effects on political outcomes. However, there is much scope for future work. First, it is important to get a better understanding of the channels of causation. The model suggests that the internet influenced elections via the media market. In line with the model's predictions, anecdotal evidence suggests a drop in the popularity of BN owned newspapers and even a decrease in bias among some media outlets as a means to reestablish credibility. Finally, it would be fruitful to explore the internet's consequences in terms of economic development. Malaysia invested heavily in internet infrastructure to promote an information economy. An important next step would be to gauge whether this investment paid off, as it would imply a relationship between political openness and economic growth.

REFERENCES

Abbott, J. (2011): "Electoral Authoritarianism and the Print Media in Malaysia: Measuring Political Bias and Analyzing Its Cause," *Asian Affairs: An American Review*, 38(1), 1–38.

Banerjee, A., S. Kumar, R. Pande, and F. Su (2010): "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India," .

Barham, T., M. Lipscomb, and A. Mobarak (2011): "Development Effects of Electrification: Evidence from the Geologic Placement of Hydropower Plants in Brazil," *CEPR Discussion Paper No. DP8427.*

Baron, D. P. (2006): "Persistent media bias," *Journal of Public Economics*, 1845(1845), 1–36.

Besley, T., and R. Burgess (2002): "The political economy of government responsiveness: Theory and evidence from india," *Quarterly Journal of Economics*, 117(4), 1415–1451.

Besley, T., and A. Prat (2006): "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability," *American Economic Review*, 96(3), 720–736.

Burchfield, M., H. Overman, D. Puga, and M. Turner (2006): "Causes of Sprawl: A Portrait from Space," *The Quarterly Journal of Economics*, 121(2), 587–633.

Case, W. (2001): "Malaysia's Resilient Pseudodemocracy," *Journal of Democracy*, 12(1), 43–57.

Centre for Independent Journalism Malaysia (2008): "Report on the Quantitative Analysis of the Media Monitoring Initiative for the 12th General Elections," pp. 1–41.

DellaVigna, S., and E. Kaplan (2007): "The fox news effect: Media bias and voting," *The Quarterly Journal of Economics*, 122(3), 1187–1234.

Dinkelman, T. (2012): "The effects of rural electrification on employment: New evidence from South Africa," *The American Economic Review, (forthcoming).*

Duflo, E., and R. Pande (2007): "Dams," *Quarterly Journal of Economics*, 122(2), 601–646.

Edmond, C. (2011): "Information manipulation, coordination and regime change," .

Gentzkow, M., and J. M. Shapiro (2011): "Ideological Segregation Online and Offline," *The Quarterly Journal of Economics*, 127(4), 1–49.

Goyal, A. (2010): "Information, Direct Access to Farmers, and Rural Market Performance in Central India," *American Economic Journal: Applied Economics*, 2(3), 22–45.

Hai, L. H. (2002): "Electoral Politics in Malaysia: 'managing' Elections in a Plural Society," in *Electoral Politics in Southeast and East Asia*, ed. by A. Croissant, G. Bruns, and M. John, pp. 101–148. Friedrich Ebert Stiftung, Singapore.

Jack, W., and T. Suri (2011): "Risk Sharing and Transaction Costs: Evidence from Kenya's Mobile Money Revolution," *MIT.*

Jensen, R. (2007): "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector," *The Quarterly Journal of Economics*, 122(3), 879–924.

Kua, K. S. (ed.) (1990): *Mediawatch: the use and abuse of the Malaysian Press.* Resource & Research Centre, Selangor Chinese Assembly Hall, Kuala Lumpur.

Mullainathan, S., and A. Shleifer (2005): "The Market for News," *American Economic Review*, 95(4), 1031–1053.

Nunn, N. (2008): "The Long-Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics*, 123(1), 139–176.

Nunn, N., and D. Puga (2009): "Ruggedness: The blessing of bad geography in Africa," *NBER Working Paper No. 14918*, pp. 1–31.

Olken, B. A. (2009): "Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages," *American Economic Journal: Applied Economics*, 1(4), 1–33.

Paluck, E. L. (2009): "Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda," *Journal of personality and social psychology*, 96(3), 574–87.

Pepinsky, T. B. (2007): "Malaysia: Turnover Without Change," *Journal of Democracy*, 18(1), 113–127.

Reinikka, R., and J. Svensson (2011): "The power of information in public services: Evidence from education in Uganda," *Journal of Public Economics*, 95(7-8), 956–966.

Salazar, L. C. (2007): *Getting a Dial Tone*. Institute of Southeast Asian Studies, Singapore.

Snyder, J., and D. Strömberg (2010): "Press Coverage and Political Accountability," *Journal of Political Economy*, 118(2), 355–408.

Strömberg, D. (2004): "Mass media competition, political competition, and public policy," *The Review of Economic Studies*, 71(1), 265.

Yanagizawa-Drott, D. (2010): "Propaganda and Conflict : Theory and Evidence From the Rwandan Genocide," .

Zinnbauer, D. (2003): "Power and Activism in the Context of a Maturing Internet : The Case of Malaysia," Ph.D. thesis, London School of Economics and Political Science.

9. TABLES

TABLE I

Summary Statistics

Variable	Mean	Std. Dev.	N
Dependent Variables			
$\Delta BNShare$ 2004-2008	-.1211	.0933	439
$\Delta Turnout$ 2004-2008	.0189	.0360	439
Independent Variables			
% Internet 2000	.1657	.1619	439
% Malay 2004	.6339	.2752	439
% Indian 2004	.07676	.0774	439
% Internet 2004	.1677	.1745	439
GDP per capita 2005	16668.72	7141.14	439
Eligible Voters 2004	17716.52	7158.29	439
Population Density	790.73	1404.24	439
% Urban	.2148	.2390	439
% Rural	.5022	.2501	439
Slope Std. Dev.	4.030	2.953	439
Road Density	.6153	.6272	439
Km to Federal Road	3.575	4.727	439
Km to Major Road	1.381	2.134	439
Instrumental Variables			
Km to Time	15.349	18.468	439
Km to Fiberail	22.400	28.184	439
Km to Fiberail Sq	1294.351	3080.338	439
Km to TM	7.129	7.787	439

Notes. The table reports summary statistics for state legislature districts in peninsular Malaysia, excluding Kuala Lumpur. Variables measured in 2008 unless otherwise stated. See appendix for details on the construction and sources of variables.

TABLE II

Evaluation of internet penetration measures

	% households with internet 2004				
	(1)	(2)	(3)	(4)	(5)
IPSumPerVoter 2005	0.360	0.580	0.265	0.621	0.628
IPMaxPerVoter 2005	0.463	0.511	0.373	0.533	0.539
IPAvgPerVoter 2005	0.317	0.459	0.169	0.492	0.515
IPFixPerVoter 2005	-0.055	-0.105	-0.584	-0.135	-0.133
IPSumPerVoter 2004	0.053	0.016	0.085	0.006	0.017
IPMaxPerVoter 2004	0.068	0.033	0.159	0.022	0.034
IPAvgPerVoter 2004	0.058	0.020	0.078	0.010	0.021
IPFixPerVoter 2004	-0.080	-0.128	-0.144	-0.153	-0.152
Kuala Lumpur	Y	N	N	N	N
Sabah	Y	Y	Y	N	N
Peninsular Malaysia	Y	Y	N	Y	Y
N	518	505	60	445	

Notes. Correlation between percentage households with internet subscription 2004 and self-constructed internet penetration measures. Percentage Households with internet access in 2004 was derived from Household Basic Amenities Survey 2004. See Section 5 for details on the construction and source of variables.

TABLE III

Relationship between BN share and internet growth from 2004 to 2008

	Dependent variable is $\Delta BNShare$ 2004-2008			
	(1)	(2)	(3)	(4)
IPperVoter growth	-0.019***	-0.013***	-0.009***	-0.009***
	(0.004)	(0.003)	(0.003)	(0.003)
% Malay 2004		0.164***	0.133***	0.130***
		(0.016)	(0.018)	(0.019)
% Indian 2004		-0.348***	-0.372***	-0.381***
		(0.053)	(0.052)	(0.054)
GDP per capita			-0.036***	-0.034***
			(0.009)	(0.010)
log Eligible Voters 2004				0.003
				(0.011)
Population Density				-0.002
				(0.003)
N	427	427	427	427
R^2	.441	.695	.71	.711

Notes. The table reports OLS estimates of equation (3). All specifications include 11 state trends. IPperVoter growth is natural log of IP addresses per eligible voter in 2008 divided by IP addresses per voter in 2004. See appendix for details on the construction and sources of variables. Coefficients are reported with robust standard errors in brackets. ***,**, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE IV

Relationship between BN share and internet growth from 1995 to 1999

	Dependent variable is $\Delta BNShare$ 1995-1999			
	(1)	(2)	(3)	(4)
InternetHH 1995-1999	0.021***	-0.015***	-0.018***	-0.020***
	(0.005)	(0.004)	(0.005)	(0.005)
% Malay 1999		-0.299***	-0.294***	-0.285***
		(0.022)	(0.023)	(0.023)
% Indian 1999		-0.212***	-0.206***	-0.198***
		(0.063)	(0.064)	(0.064)
GDP per capita 2005			0.012	0.008
			(0.011)	(0.011)
log Eligible Voters 1995				0.012
				(0.017)
Population Density 2008				0.001
				(0.001)
N	374	374	374	374
R^2	.269	.576	.577	.579

Notes. The table reports OLS estimates of equation (3). All specifications include 11 state trends. 1999 election is from December 1999. Internet growth is the natural log percentage of households with internet subscriptions in 2000 (internet access was zero in 1995). See appendix for details on the construction and sources of variables. Coefficients are reported with robust standard errors in brackets. ***,**, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE V

First Stage relationship between distance to backbone and internet growth

	Growth in IPs per eligible voter 2004-2008: $\triangle IPperVoter$					
	(1)	(2)	(3)	(4)	(5)	(6)
Km to Time*10	-0.115***	-0.092***	-0.094***	-0.082***	-0.082***	-0.082***
	(0.027)	(0.027)	(0.028)	(0.029)	(0.029)	(0.029)
Km to Fiberail*10	-0.194***	-0.177***	-0.177***	-0.158***	-0.156***	-0.157***
	(0.043)	(0.043)	(0.044)	(0.045)	(0.045)	(0.045)
Km to Fiberail*10 SQ	0.021***	0.019***	0.019***	0.018***	0.018***	0.018***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Km to TM*10	-0.015	-0.014	-0.024	0.003	-0.002	-0.002
	(0.059)	(0.058)	(0.058)	(0.064)	(0.064)	(0.064)
% Malay 2004	-0.001	0.295	0.173	0.297	0.294	0.303
	(0.196)	(0.225)	(0.238)	(0.240)	(0.239)	(0.259)
% Indian 2004	0.134	0.405	0.093	0.388	0.457	0.483
	(0.648)	(0.661)	(0.665)	(0.682)	(0.684)	(0.725)
GDP per Capita		0.384***	0.465***	0.390***	0.402***	0.404***
		(0.116)	(0.122)	(0.124)	(0.124)	(0.127)
log Eligible Voters 2004			0.001	-0.035	-0.035	-0.039
			(0.157)	(0.155)	(0.155)	(0.162)
Population Density			-0.077**	-0.148***	-0.133***	-0.133***
			(0.036)	(0.051)	(0.051)	(0.051)
% Urban				0.770**	0.965***	0.962***
				(0.335)	(0.352)	(0.351)
% Rural				-0.033	-0.018	-0.019
				(0.243)	(0.243)	(0.243)
Slope std				-0.001	-0.001	-0.001
				(0.019)	(0.019)	(0.019)
Road density					-0.131	-0.129
					(0.086)	(0.086)
BN Share 2004						-0.056
						(0.495)
N	427	427	427	427	427	427
R^2	.307	.325	.333	.349	.351	.351

Notes. The table presents OLS estimates of equation (5). It presents first stage results for the relationship between distance to backbone and growth in IP addresses per voter. All specifications include 11 state trends. All specifications include state trends. IPperVoter is natural log of IP addresses per eligible voter in 2008 divided by IP addresses per voter in 2004. Coefficients are reported with robust standard errors in brackets. ***,**, and * indicate significance at the 1%, 5%, and 10% levels. See appendix for details on the construction and source variables.

TABLE VI

IV Estimates of the relationship between BN share and internet growth

	Dependent variable is $\Delta BNShare$ 2004-2008					
			IV			OLS
	(1)	(2)	(3)	(4)	(5)	(6)
IPperVoter Growth	-0.036***	-0.028***	-0.029***	-0.037***	-0.036***	-0.009***
	(0.007)	(0.008)	(0.009)	(0.011)	(0.010)	(0.003)
% Malay 2004	0.156***	0.137***	0.133***	0.129***	0.128***	0.120***
	(0.017)	(0.018)	(0.019)	(0.020)	(0.020)	(0.019)
% Indian 2004	-0.331***	-0.350***	-0.362***	-0.374***	-0.371***	-0.396***
	(0.055)	(0.053)	(0.055)	(0.055)	(0.054)	(0.053)
GDP per capita		-0.025**	-0.022**	-0.018*	-0.017	-0.031***
		(0.010)	(0.011)	(0.011)	(0.011)	(0.010)
log Eligible Voters 2004			0.009	0.007	0.007	0.003
			(0.012)	(0.012)	(0.012)	(0.011)
Population Density			-0.004	-0.006*	-0.005	-0.000
			(0.003)	(0.003)	(0.003)	(0.003)
% Urban				-0.008	0.004	-0.028
				(0.025)	(0.029)	(0.026)
% Rural				-0.009	-0.009	-0.012
				(0.016)	(0.016)	(0.016)
Slope std				-0.003**	-0.003**	-0.003***
				(0.001)	(0.001)	(0.001)
Road Density					-0.008	-0.006
					(0.010)	(0.010)
N	427	427	427	427	427	427
R^2	.657	.686	.685	.67	.673	.717
F-Stat	18.1	13.1	12.9	8.5	8.7	
Hansen Test (p-value)	.75	.76	.73	.80	.82	

Notes. Specifications (1) through (5) show results of IV regressions of change in BN vote share 2004-2008 on IPperVoter growth 2004-2008. Instruments are distance to Time, distance to Fiberail, distance to Fiberail squared, and distance to TM. Column (6) reports results from an ordinary least squares regression of BN vote share 2004-2008 on IPperVoter growth 2004-2008. F-stat is the f-statistic of the instruments from the first stage. The *p*-value for the Hansen test is for the Sargan-Hansen test of overidentifying restrictions. The joint null is that the instruments are uncorrelated with the error. All specifications include 11 state trends. IPperVoter growth is the natural log of IP addresses per eligible voter in 2008 divided by IP addresses per voter in 2004. See appendix for details on the construction and sources of variables. Coefficients are reported with robust standard errors in brackets. ***,**, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE VII

Reduced form estimates of distance to backbone on elections

	Dependent variable is $\Delta BNShare$					
	1986-1990		1995-1999		2004-2008	
	(1)	(2)	(3)	(4)	(5)	(6)
Km to Time*100	-0.036		0.038*		0.041**	
	(0.026)		(0.023)		(0.016)	
Km to Fiberail*100		-0.055		0.040		0.035
		(0.047)		(0.038)		(0.030)
Km to Fiberail*100 SQ		0.033		-0.034		-0.057**
		(0.035)		(0.031)		(0.023)
Road Density	0.257**	0.269**	0.075	0.061	-0.004	-0.003
	(0.120)	(0.120)	(0.104)	(0.105)	(0.010)	(0.010)
Fiberail joint significance		.72		.55		.001
N	325	325	368	368	427	427
R^2	.507	.507	.567	.565	.715	.716

Notes. Reduced form regressions of change in BN share on distance to the backbone are reported. Columns (1) and (2) cover the 1986-1990 elections; columns (3) and (4) cover the 1995-1999 elections; and columns (5) and (6) cover the 2004-2008 elections. Fiberail joint significance presents the p-value of a test of the joint significance of Km to Fiberail and Km to Fiberail squared. All specifications control for ethnicity, GDP per capita, percent of the district that is urban and rural, the log of eligible voters, population density and 11 state trends. GDP per capita is taken from a 2005 estimate in all cases. For all specifications population density, road density, % urban, and % rural are calculated from a 2008 measure. Coefficients are reported with robust standard errors in brackets. ***,**, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE VIII

IV Estimates controlling for distance to roads

	Dependent variable is $\Delta BNShare$ 2004-2008								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IPperVoter growth	-0.036***	-0.033***	-0.032***	-0.035***	-0.033***	-0.032***	-0.040**	-0.035**	-0.035**
	(0.010)	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)	(0.017)	(0.017)	(0.017)
Km to Major Road*10		0.017*			0.017*			0.017*	
		(0.009)			(0.009)			(0.010)	
Km to Federal Road*10			0.010*			0.011*			0.010*
			(0.005)			(0.005)			(0.006)
N	427	427	427	427	427	427	427	427	427
R^2	.673	.684	.686	.677	.684	.686	.659	.678	.679
F-stat	8.7	9.2	9	14.5	14.4	14.3	12.5	13	12.8
Instrumental Variables	ALL	ALL	ALL	Fiberail	Fiberail	Fiberail	Time	Time	Time

Notes. Specifications (1) through (9) show results of IV regressions of change in BN vote share 2004-2008 on IPperVoter growth 2004-2008. Instruments in (1)-(3) are distance to Time, distance to Fiberail and Fiberail squared, and distance to TM. Instruments in (4)-(6) are distance to Fiberail and distance to Fiberail squared. The instrument in (7)-(9) is distance to Time. F-stat is the f-statistic of the instruments from the first stage. All specifications control for ethnicity, GDP per capita, percent of the district that is urban and rural, the log of eligible voters, population density and 11 state trends. IPperVoter growth is natural log of IP addresses per eligible voter in 2008 divided by IP addresses per voter in 2004. See appendix for details on the construction and sources of variables. Coefficients are reported with robust standard errors in brackets. ***,**, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE IX

Placebo regressions of $\Delta BNShare$ on internet growth in a different time period

	(1)	(2)	(3)	(4)	(5)
PANEL A: Dependent variable is $\Delta BNShare$ 1986-1990					
IPperVoter 2004-2008	-0.002	-0.002	-0.002	-0.004	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
PANEL B: Dependent variable is $\Delta BNShare$ 1995-1999					
IPperVoter 2004-2008	-0.001	-0.005	-0.005	-0.005	-0.005
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
Controls					
Ethnicity	N	Y	Y	Y	Y
GDP per capita	N	N	Y	Y	Y
Population	N	N	N	Y	Y
Road	N	N	N	N	Y

Notes. The table reports OLS estimates of equation BN share change on internet growth in a different period. Panel A reports results of BN share 1986-1990 on IPperVoter Growth from 2004 to 2008. Panel B reports results of BN share 1995-1999 on IPperVoter Growth from 2004 to 2008. IPperVoter 2004-2008 is the natural log of IP addresses per eligible voter in 2008 divided by IP addresses per voter in 2004. All specifications include 11 state trends. Ethnicity controls are % Malay and % Chinese, from each respective period. GDP per capita is taken from 2005 for panels A and B. Population controls for population density, road density, % urban, % rural, and log of eligible voters. Log of eligible voters is for 1986, 1995, and 2004, respectively, while the other controls are from 2008. Road controls for road density and distance to federal roads as of 2008. See appendix for details on the construction and sources of variables. Coefficients are reported with robust standard errors in brackets. ***,**, and * indicate significance at the 1%, 5%, and 10% levels. For expository clarity, coefficients on controls are not reported.

TABLE X

Probit estimates of turnover on internet

	Probit		IV Probit			
	1999	2008	2008	2008	2008	2008
	(1)	(2)	(3)	(4)	(5)	(6)
IPperVoter 2008		-0.286**	-0.946*	-1.041**	-1.109	-1.011
		(0.140)	(0.557)	(0.513)	(0.702)	(0.617)
InternetHH 1999	0.300					
	(0.222)					
N	255	383	383	383	383	383
Pseudo R^2	.436	.51				
First stage: Dependent variable is $IPperVoter$ 2008						
Km to Time*10			-0.044	-0.040	-0.050*	
			(0.027)	(0.027)	(0.028)	
Km to Fiberail*10			-0.007	-0.001		-0.005
			(0.044)	(0.042)		(0.043)
Km to Fiberail*10 SQ			0.005	0.004		0.005
			(0.004)	(0.004)		(0.004)
Km to TM*10			0.070			
			(0.059)			
F-stat			2.7	3.28	3.47	3.5
N			383	383	383	383

Notes. Probit estimates of turnover on internet connectivity are reported. Specification (1) regresses turnover from December 1999 on log % households with internet subscription in 1999, and restricts sample to districts won by the BN in 1995. Specifications (2)-(6) regress turnover 2008 on log IPperVoter 2008, and restrict sample to districts that the BN won in 2004. All specifications control for ethnicity, GDP per capita, percent of the district that is urban and rural, the log of eligible voters, population density, road density, distance to federal roads, and 11 state trends. GDP per capita is taken from a 2005 estimate. For all specifications distance to federal roads, road density, % urban, and % rural are calculated from a 2008 measure. F-stat is the f-statistic of the instruments from the first stage of 2SLS estimate from the equivalent linear probability model. See appendix for details on the construction and sources of variables. Coefficients are reported with standard errors in brackets. ***,**, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE XI

Relationship between turnout and internet growth

	$\Delta Turnout$ 2004-2008				$\Delta Turnout$ 1995-1999
	OLS	OLS	IV	IV	OLS
	(1)	(2)	(3)	(4)	(5)
IPperVoter growth 04-08	0.0035*	0.0042**	0.0138	0.0157*	
	(0.0019)	(0.0019)	(0.0087)	(0.0085)	
$InternetHH$ 1995-1999					0.0034
					(0.0026)
Drop irregularities	N	Y	N	Y	N
Time IV	N	N	Y	Y	N
N	427	413	427	413	368
R^2	.649	.641	.604	.586	.301

Notes. Specifications (1) and (2) show results of OLS regressions of change in turnout 2004-2008 on IPperVoter growth 2004-2008. Columns (3) and (4) present results of IV regressions using distance to Time as an instrument. Specification (5) reports results of regressions of change in turnout 1995-1999 on internet subscription per household growth 1995-1999. Drop irregularities drops districts with irregularities in turnout; see appendix for details. InternetHH 1995-1999 is the natural log percentage of households with internet subscriptions in 2000 (internet access was zero in 1995). All specifications control for ethnicity, GDP per capita, percent of the district that is urban and rural, the log of eligible voters, population density, road density and 11 state trends. For specifications (5) GDP per capita is taken from 2005. Road density, % urban, % rural are from 2008. IPperVoter growth is natural log of IP addresses per eligible voter in 2008 divided by IP addresses per voter in 2004. See appendix for details on the construction and sources of variables. Coefficients are reported with robust standard errors in brackets. ***,**, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE XII

Results of state legislature elections without internet

State	$BNSeats2004$	$BNSeats2008$	$\widehat{BNSeats}2008_{OLS}$	$\widehat{BNSeats}2008_{IV}$
Johor	.982	.892	.946	.946
Kedah	**.861**	**.388**	**.361**	**.667**
Kelantan	.466	.133	.311	.489
Melaka	.928	.821	.857	.857
Negeri Sembilan	.944	.583	.75	.75
Pahang	.976	.880	.928	.952
Perak	**.881**	**.474**	**.576**	**.644**
Perlis	.933	.933	1	1
Pulau Pinang	**.95**	**.275**	**.35**	**.45**
Selangor	**.964**	**.357**	**.392**	**.554**
Terengganu	.875	.75	.812	.937
N	445	445	445	445

Notes. Table reports fraction of state legislature seats won by the BN alongside estimates in the absence of internet. Covers all state peninsular seats.$BNSeats2004$ and $BNSeats2008$ are the fraction of state legislature seats won by the BN in 2004 and 2008 respectively. $\widehat{BNSeats}2008_{OLS}$ is the predicted fraction of seats won by the BN in the absence of internet growth based on OLS equation (3). $\widehat{BNSeats}2008_{IV}$ is the predicted fraction of seats won by the BN in the absence of internet growth based on the IV system of equations (4) and (5).

10. FIGURES

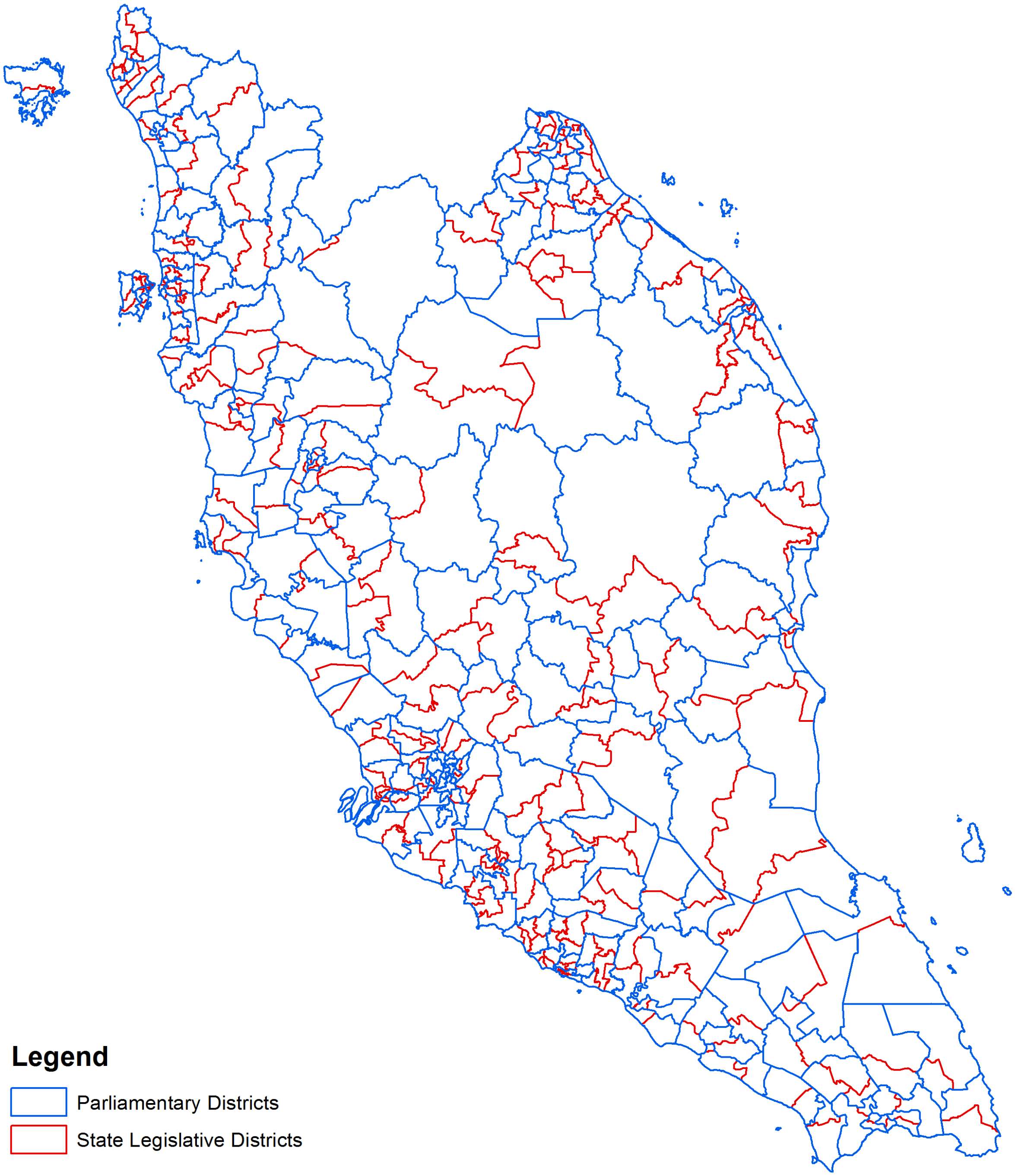

FIGURE 1.— Political Boundaries

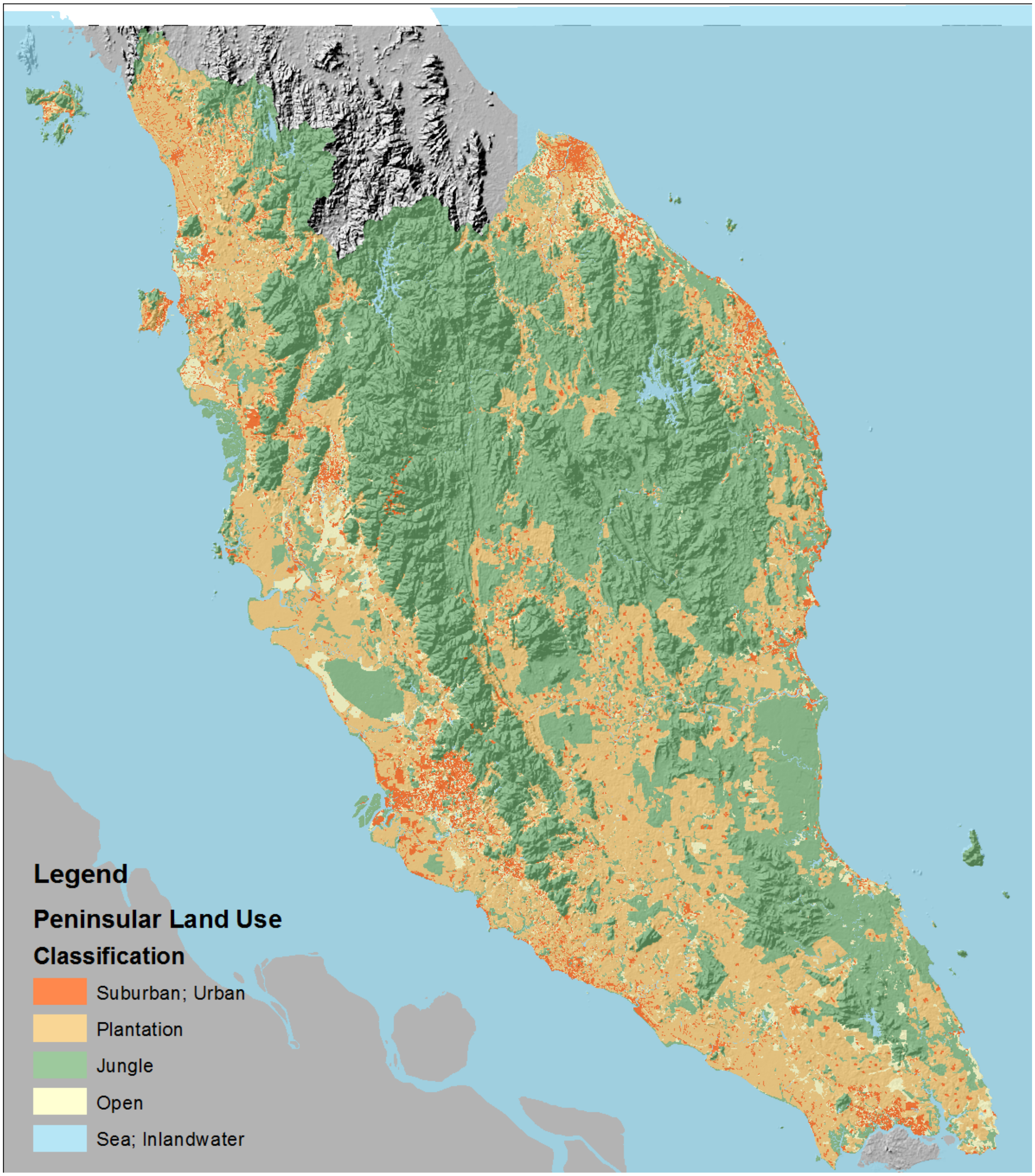

Figure 2.— Peninsular Malaysia Land Use and Elevation

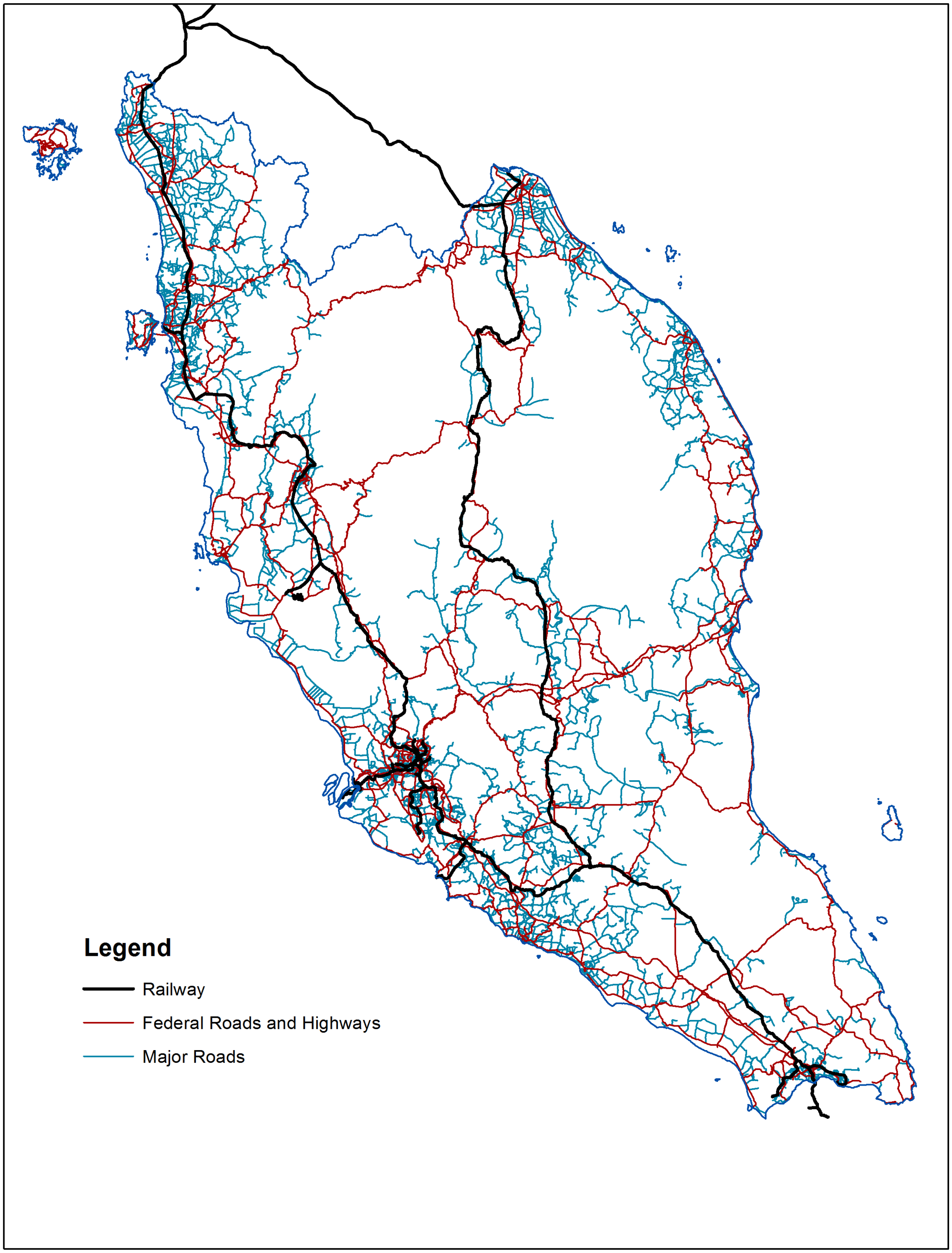

FIGURE 3.— Peninsular Malaysia's road and railway network

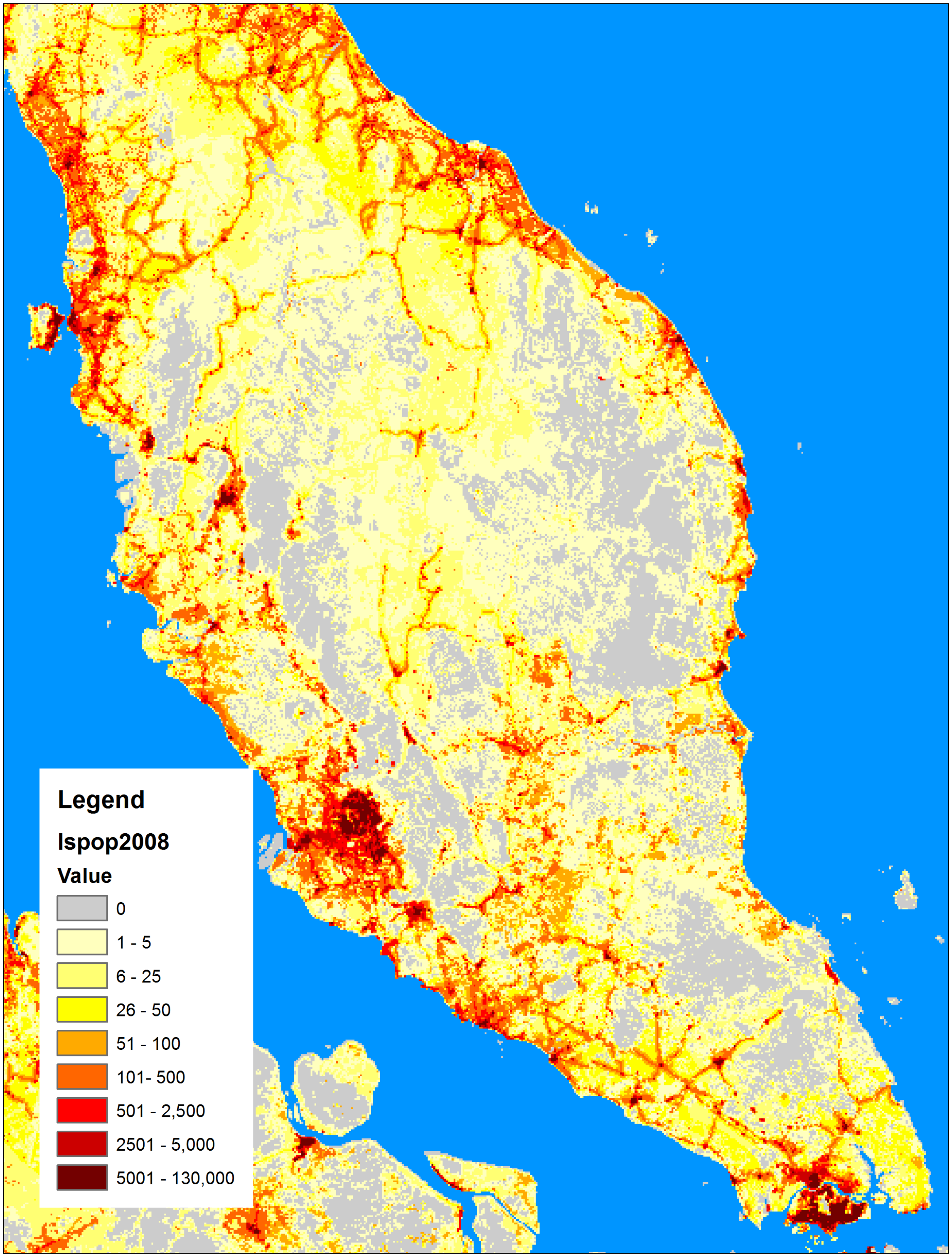

Figure 4.— Population Distribution

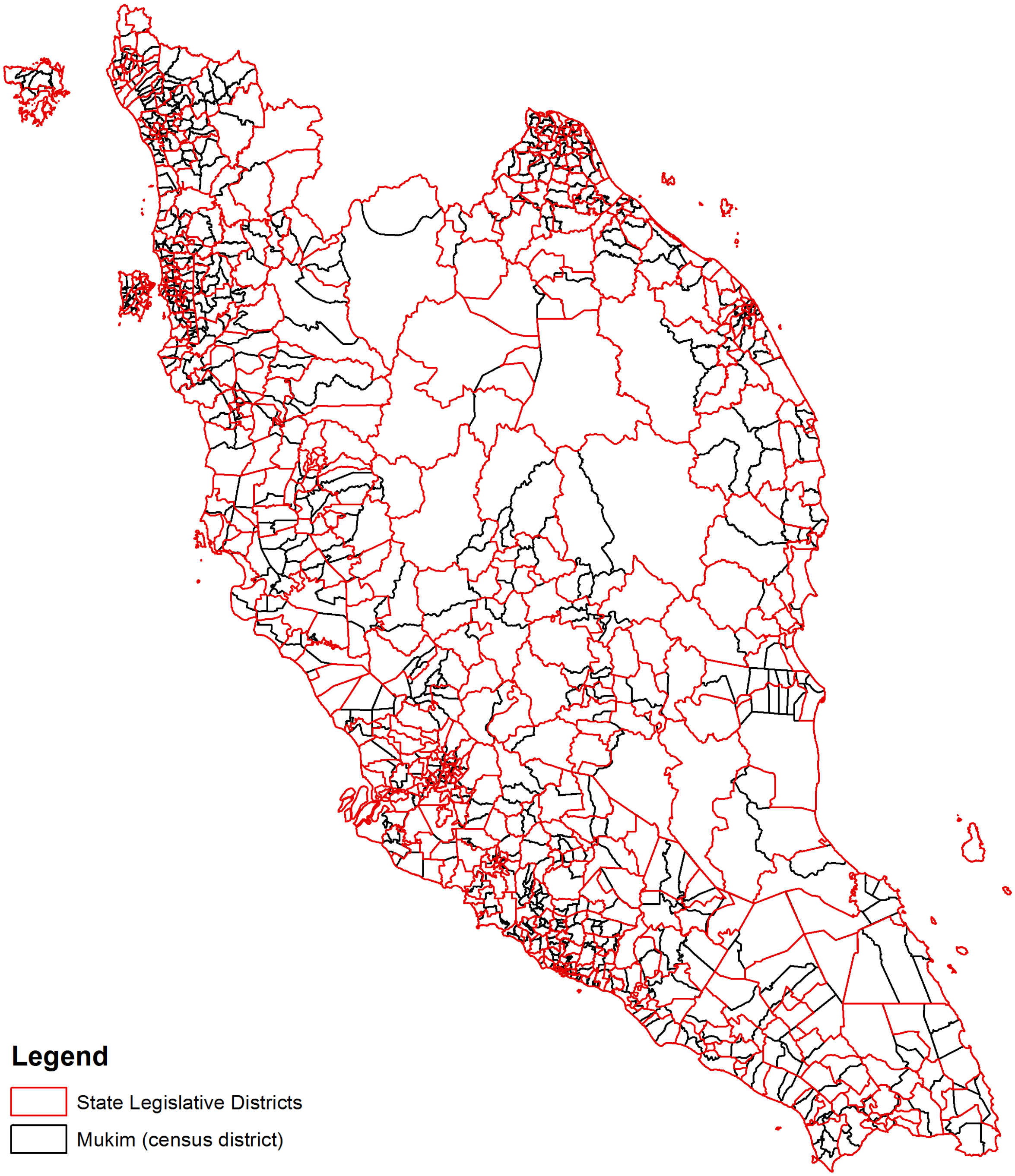

Figure 5.— Legislative District vs. Census District Boundaries

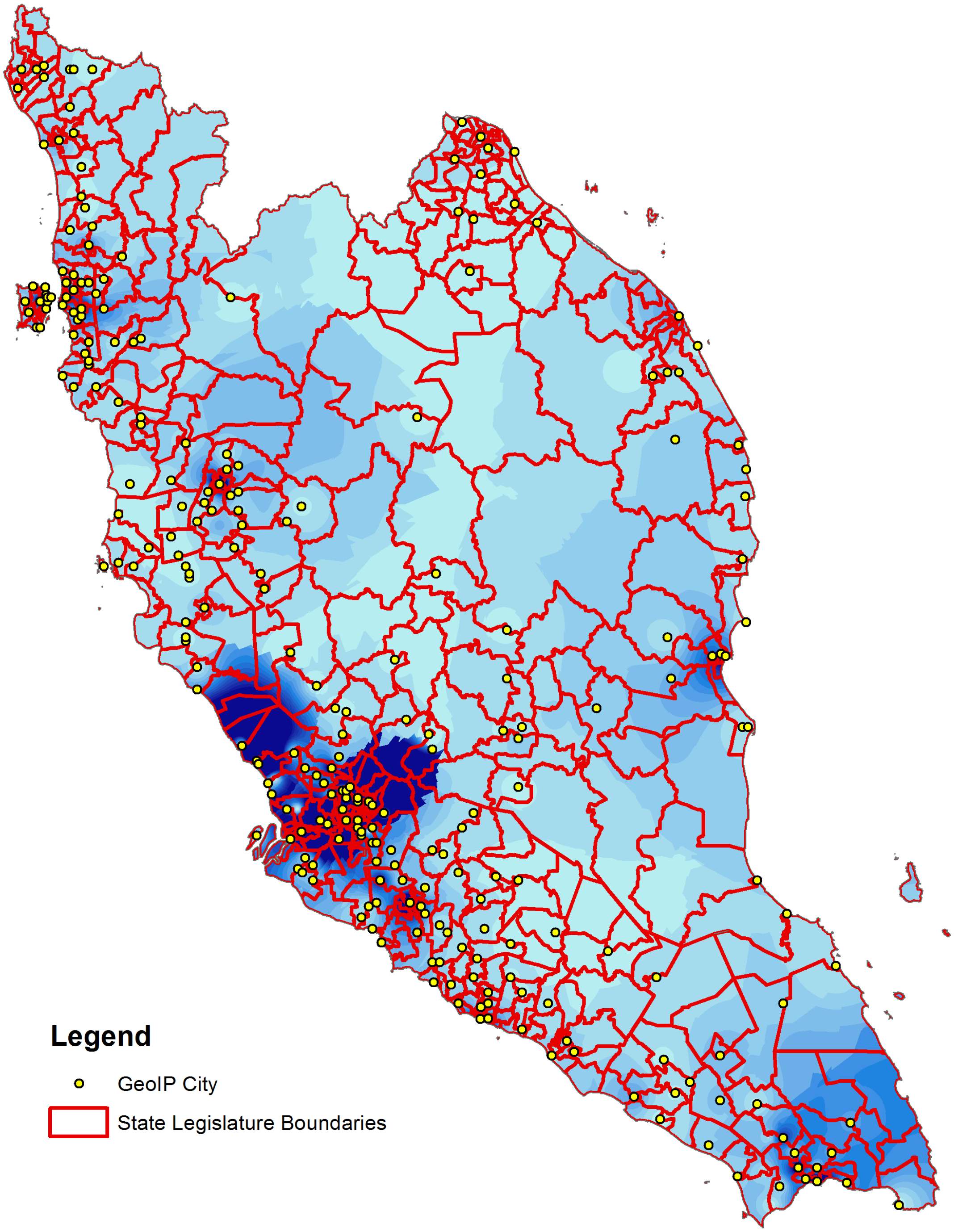

Figure 6.— Inverse Distance Weighting Interpolation

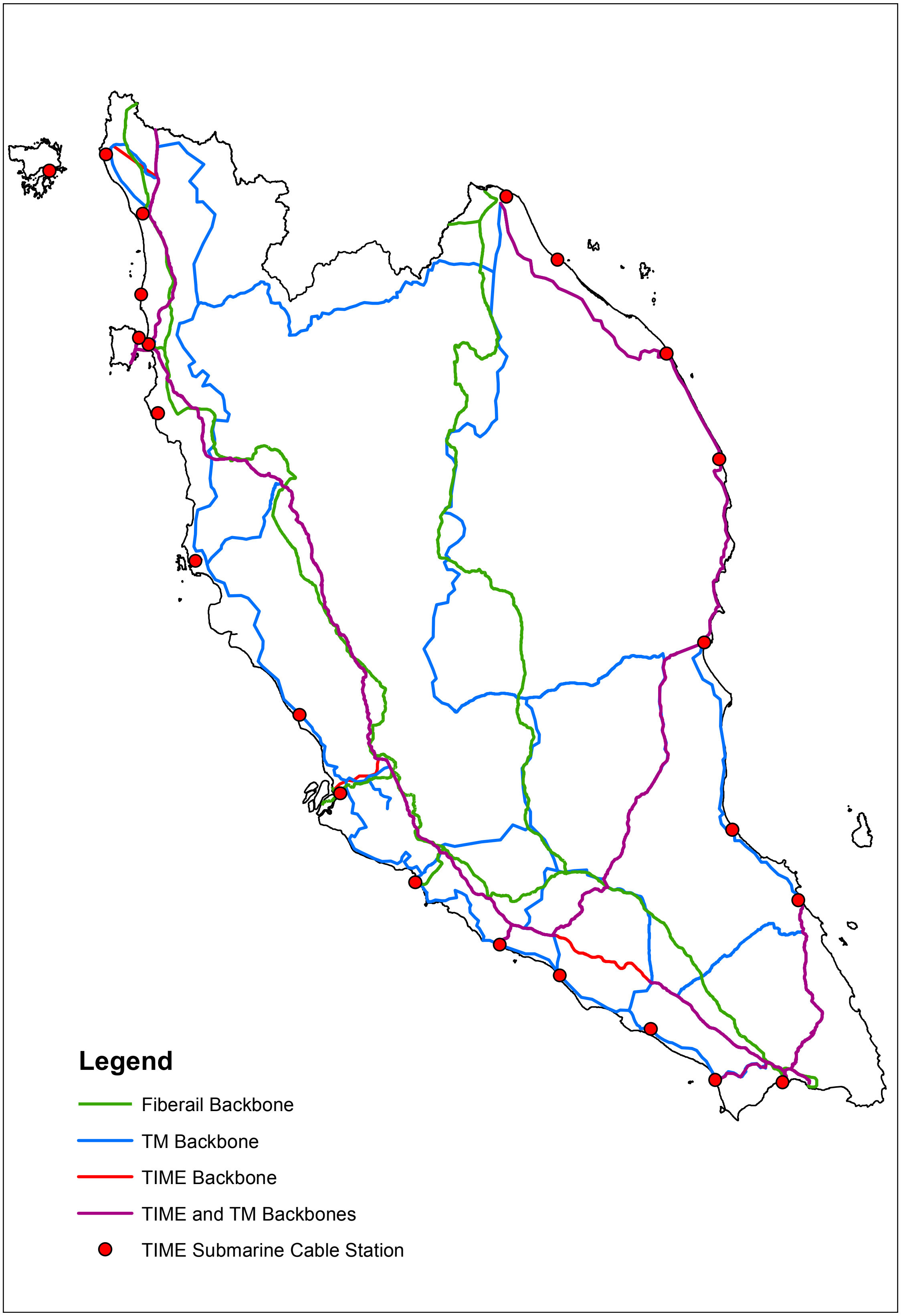

FIGURE 7.— Backbone Location

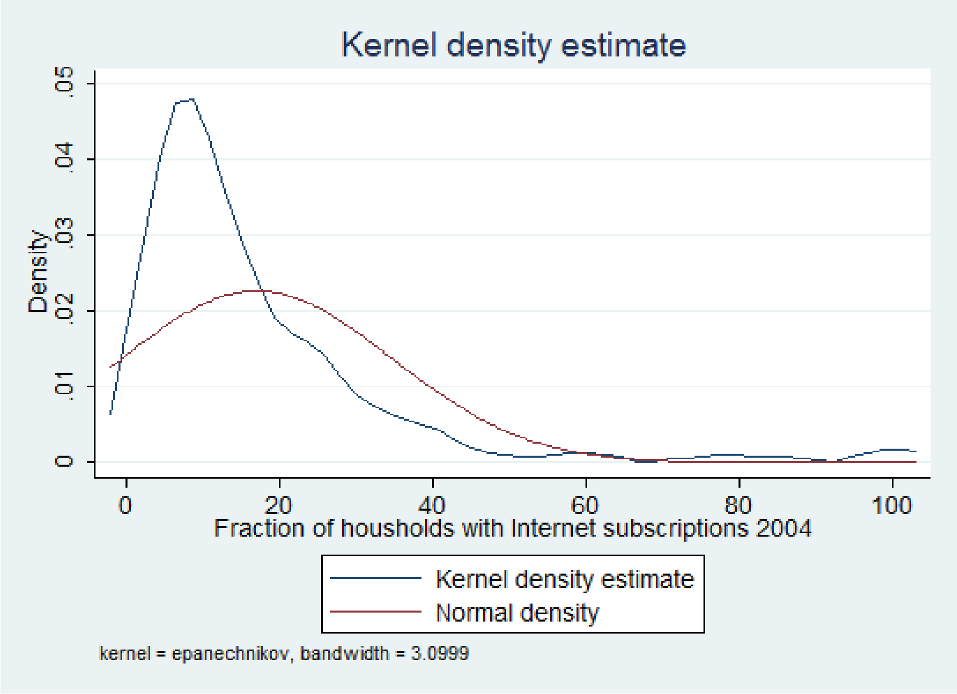

FIGURE 8.— Skewness of households with internet 2004

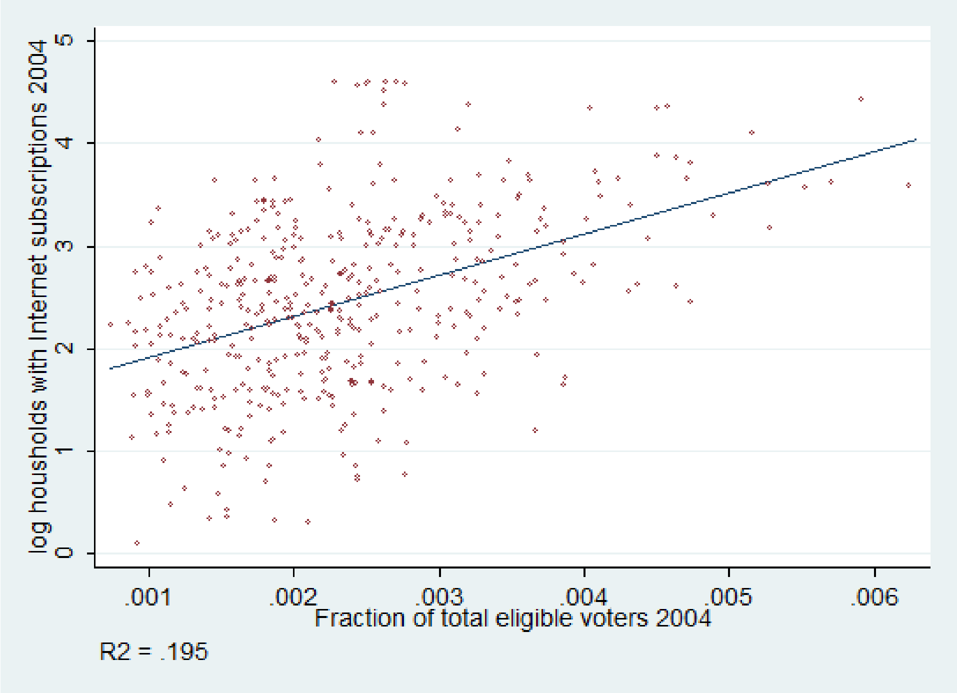

FIGURE 9.— Relationship between log households with internet 2004 and fraction of total eligible voters

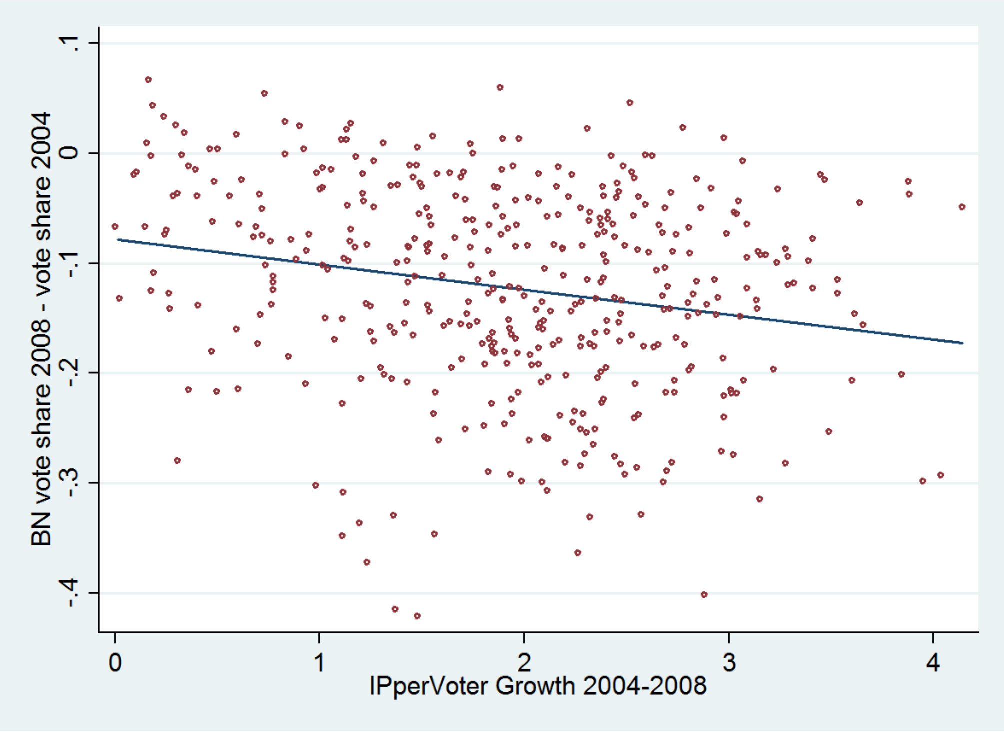

Figure 10.— Relationship between change in BN vote share and IP per voter growth

APPENDIX

APPENDIX A: DERIVATION OF BIAS FROM THE UNDERSAMPLING OF INTERNET CONNECTIVITY IN AREAS FAR FROM LARGE CITIES

In this section I draw from Nunn (2008) to show that the undersampling of IP addresses per voter for areas outside the largest cities will result in OLS estimates of the effect of IP per voter growth on incumbent vote share that are biased toward zero.

Denote the true log IP addresses per voter in district i as s^*_{it}, the observed number as s_{it}, the distance to the nearest large city d_i and incumbent vote share by y_{it}. d_i is expressed as deviation from the mean. Assume the true relationship between the change in log of IP addresses and distance from a major city is:

$$\text{(A.1)} \quad \triangle s^*_{it} = -\alpha d_i + \triangle e_{it}$$

where $\alpha > 0$ is and e_{it} is i.i.d. and drawn from a normal distribution.

Next, turning to the undersampling of regions farther from large cities, assume that the relationship between the observed change in the log of IP addresses, $\triangle s_{it}$, and the distance to the nearest major city is given by:

$$\text{(A.2)} \quad \triangle s_{it} = \triangle s^*_{it} - \gamma d_i + \triangle v_{it}$$

where $\gamma > 0$ and v_{it} is uncorrelated with e_{it} and d_i.

The true relationship between change in BN share and change in log IP addresses per voter is given by

$$\text{(A.3)} \quad \triangle y_{it} = -\beta \triangle s^*_{it} + \triangle w_{it}$$

where $\beta > 0$ and w_{it} is uncorrelated with all other variables.

If we perform a simple OLS estimate of the effect of observed IP per voter on BN share,

$\triangle y_{it} = b\triangle s_{it} + \triangle u_{it}$, we get:

$$\text{(A.4)} \quad \hat{b} = \frac{\sum_i \triangle s_{it} \triangle y_{it}}{\sum_i \left(\triangle s_{it}\right)^2}$$

Substituting (A.1) into (A.2) and taking the first difference gives:

$$\text{(A.5)} \quad \triangle s_{it} = -(\alpha + \gamma) d_i + \triangle e_{it} + \triangle v_{it}$$

Substituting (A.1) into (A.3) gives:

$$\text{(A.6)} \quad \triangle y_{it} = \beta \alpha d_i - \beta \triangle e_{it} + \triangle w_{it}$$

Finally substituting (A.5) and (A.6) into (A.4) and taking the plim gives :

$$\text{(A.7)} \quad plim\hat{b} = -\beta \frac{\sigma^2_{\triangle s^*} + \alpha\gamma\sigma^2_d}{\sigma^2_{\triangle s^*} + \gamma(2\alpha + \gamma)\sigma^2_d + \sigma^2_{\triangle \nu}}$$

where $\sigma^2_{\triangle s^*} = \alpha^2 \sigma^2_d + \sigma^2_{\triangle e}$

In the case where $\gamma = 0$ we are in the situation of classical measurement error and the result reduces to standard attenuation bias.

If $\gamma > 0$ we are in a situation where the underestimation of IP addresses per voter is increasing in distance from a major city. Since $2\gamma(\alpha + \gamma) > \alpha\gamma$, (A.7) will be biased toward zero in this case as well.

APPENDIX B: PROOFS OF PROPOSITIONS

B.1. *Proof of Proposition 1*

As in Besley and Prat (2006), I focus on pure strategy, perfect Bayesian equilibria.

First I start with the voters. Since voting is sincere, each voter observes media reports and updates the posterior probability that the incumbent is good $\hat{\gamma}$. She votes for the incumbent if and only if $\hat{\gamma} > \gamma$. Suppose that viewers observe each of the two signal realizations with positive probability. Then it must be that $\hat{\gamma}(\tilde{s}_i = b) < \gamma < \hat{\gamma}(\tilde{s}_i = \emptyset)$. To see this, suppose there is a pure strategy equilibrium where the incumbent is kicked out if $s = \emptyset$: $\hat{\gamma}(\tilde{s}_i = \emptyset) < \gamma$. If this were the case, then the incumbent would never buy off the media. This in turn would cause voters to update their posterior such that $\hat{\gamma}(\tilde{s}_i = \emptyset) > \gamma$, a contradiction. This means that if $s = \emptyset$ the incumbent party is always reelected and if $s = b$ the incumbent party is reelected if and only if at least half of the voters observe $\tilde{s}_i = \emptyset$.

Now I move to the decision of media firms. Suppose that τ_W is so high that the incumbent will never choose to capture the web based media firm. Next, suppose that the mainstream outlet has been offered t_M to suppress its signal and it knows that the web firm will not suppress its signal. The mainstream firm's payoff is:

$$\pi_M = \begin{cases} a\left(1 - \frac{\phi}{2}\right) & \text{if she rejects} \\ \frac{t_M}{\tau_M} & \text{if she accepts} \end{cases}$$

Thus she accepts if and only if

$$t_M \geq \tau_M a \left(1 - \frac{\phi}{2}\right)$$

Finally, consider the incumbent. The web media is too costly to capture by definition. If $\Phi \geq \frac{1}{2}$ at least half of the voters will receive the bad signal and the incumbent will lose regardless of whether the traditional media is captured. Thus a bad incumbent will not capture either outlet and will be discovered with probability q. If $\Phi < \frac{1}{2}$ a bad incumbent will capture the mainstream firm if the return from reelection is greater than the cost

$$r \geq \tau_M a \left(1 - \frac{\phi}{2}\right)$$

I have thus shown that the incumbent will capture the media under the conditions specified in Proposition I.

B.2. *Proof of Proposition 2*

The effect on turnover and the media is explained in the text. The remaining results follow directly from Proposition 1.

B.3. *Proof of Proposition 3*

The proof is identical to Proposition 1 except for the equilibrium strategies and decisions of the incumbent.

Consider the incumbent. Assume $\Phi \geq \frac{1}{2}$ and either one or both of the following conditions are met, such that $\phi < \frac{1}{2}$ for a majority of districts:

1. Internet penetration is rightly skewed around $\phi = \frac{1}{2}$
2. ϕ and ψ are positively correlated

Then a bad incumbent will capture the mainstream media as the conditions ensure the victory in a majority of seats.

APPENDIX C: DATA APPENDIX

Data Descriptions and Sources

Variable	**Description**	**Source**
Original Variables		
BNShare	Share of the vote won by a member of the Barisan Nasional.	Election Commission (1986, 1990, 1995, 1999, 2004, 2008)
Turnout	Percentage of eligible voters who voted in a district. Includes spoilt ballots.	Election Commission (1986, 1990, 1995, 1999, 2004, 2008)
Turnover	Dummy indicating whether BN retained seat.	Election Commission (1999, 2008)
Eligible Voters	Number of people eligible to vote in a district.	Election Commission (1986, 1990, 1995, 1999, 2004, 2008)
% ethnicity	Percentage of the voters in a district who are of each ethnic group. The omitted category is Chinese/Other. This number is reported as part of the election results in all major dailies.	Malaysiakini (2004), New Strait Times Press (1999)
Generated Variables		
GDP	Measure of average GDP per capita at the mukim (census district) level, generated by the consultancy Booz & Company. Aggregated up to the state legislature district level using ArcGIS and LandScan as outlined in Section 4.2.	Booz & Company (2005)
InternetHH	Fraction of households with an internet subscription at the mukim level. Aggregated up to the state legislature district level using ArcGIS and LandScan as outlined in Section 4.2.	Household Basic Amenities and Income Survey (2004), Population and Housing Census (2000)

% ethnicity 1991	Percentage of the voters in a district who are of each ethnic group. The omitted category is Chinese/Other. Data is available at the mukim level in 1991 population census. Aggregated up to the state legislature district level using ArcGIS and LandScan as outlined in Section 4.2.	Population and Housing Census (1991)
Slope Std.	The standard deviation of the average steepness of land in a district. Calculated from digital elevation satellite imagery using ArcGIS first to calculate the slope at each point and then to derive the average and standard deviation across a district.	Pitney Bowes (2008)
% Urban Rural	Percentage of a district that is classified as urban and rural farm using satellite imagery. The omitted category is jungle. Used ArcGIS to calculate percentage at district level.	Pitney Bowes (2008)
Road Density	Kilometers of road in a district divided by total area of the district. Calculated using ArcGIS.	Pitney Bowes (2008)
Population Density	From Oak Ridge National Laboratory LandScan product, which uses census data in conjunction with satellite information to estimate population at the 1 km resolution. I use ArcGIS to aggregate up to the district level.	LandScan, Oak Ridge National Laboratory (2008)
Km to roads	Distance from the centroid of a district to the closest major road and closest federal road as of 2008. The road data is from Pitney Bowes. ArcGIS was used to calculate the centroid of each district and then derive the distance measure.	Pitney Bowes (2008)

Km to Time	Shortest distance from the centroid of a district to Time dotCom's backbone. Location of Time dotCom's backbone from company records. Distance generated in ArcGIS.	Time dotCom (2004)
Km to Fiberail	Shortest distance from the centroid of a district to Fiberail's backbone, which follows major railroads. Distance generated in ArcGIS.	Pitney Bowes (2008)
Km to TM	Shortest distance from the centroid of a district to Telekom Malaysia's backbone. From 2004 annul report. Distance generated in ArcGIS.	Telekom Malaysia (2004)
IPperVoter	The number of IP addresses per eligible voter. Constructed with ArcGIS from IP geolocation data from MaxMind in conjunction with records from APNIC. See Section 5 for details.	MaxMind (2004-2008), APNIC (2004-2008)

APPENDIX D: APPENDIX ON ELECTION IRREGULARITIES

There are several factors that lead to election irregularities in the 2004-2008 period. First, there were some large discrepancies between the number of ballots issued for parliamentary seats and their corresponding state legislature seats. Recall from Figure 1 that each parliamentary seat is made up of a handful of state legislature seats. Technically the number parliamentary ballots should be the same as the sum of the ballots for all the constituent state legislature seats. However, there were several large discrepancies in this respect, most notably the Kuala Terengganu parliamentary seat where there was a difference of 10,000 between the number of parliamentary ballots issued (around 70,000) and the number of state legislature ballots issued. I drop seats where the discrepancies are suspiciously large: 5 state legislature districts corresponding to the Kuala Terengganu parliamentary seat (Wakaf Mempelam, Bandar, Ladang, and Batu Buruk); and the 4 state legislature districts corresponding to the Setiu parliamentary seat (Batu Rakit, Jabi, Langkap, and Permaisuri).

The 2004 election was marked by unnaturally high turnout rates, greater than 90% in several instances. To deal with this, I drop districts where turnout exceeded 80% in 2004 and turnout differed by more than 10% from its level in the 1999 election. Due to redistricting, boundaries do not perfectly match between 1999 and 2004. In order to generate a 1999 turnout value for a 2004 district I use the population weighted LandScan procedure outlined in Section 4.2. This rule leads me to drop 6 additional districts: Lunas, Nenggiri, Sungai Udang, Chini, Kuala Nerang, and Sungai Tiang.