The Political Economy of Restructuring Eskom

Lumkile Mondi

Abstract

Eskom cannot keep lights on. This paper is about the political economy of the restructuring of Eskom from 1985 – beyond. First, the paper considers the extent to which the narrative mobilised by Eskom management has given rise to the promise of "electricity for all" through the extension of electricity to black households, both urban and rural thereby frustrating the restructuring process and the indecision by the democratic state that precipitated an energy crisis that is informing the intention to restructure Eskom. Second, the paper examines what the impact has been of the ANC macroeconomic policies advocated by black interest groups as well as the evolution of their primary legitimation mechanism, the introduction of Preferential Procurement and the capture of Eskom leading to its restructuring.

Introduction

Public choice literature (Mueller, 1976) argues that SOEs invest inefficiently relative to their private sector counterparts. The reason for this inefficiency is the efforts of political actors, unconstrained by the market forces associated with ownership and control of private sector firms, to use SOEs as a vehicle for redistributing wealth to salient political constituencies such as concentrated geographic interests, entrepreneurs, labour unions and construction firms. Therefore, in the case of South Africa, public choice scholars argue that political actors such as the National Party (NP) from 1948 to 1994 and the ANC from 1994 to date have directed SOEs to undertake white elephant investment projects that provide targeted economic benefits to their constituencies, even when the aggregate economic benefits of such projects, such as increased output or service quality, do not cover the economic (opportunity) costs borne by the broader polity. This argument has dominated the South African economic scholarship in its explanation of why apartheid was unsustainable.

In this paper I estimate Eskom's electricity capacity deployment between 1985 and 2018 by developing a machine learning (ML) framework. I then examine the effects of interest group pressure and the structure of political institutions on infrastructure deployment by state-owned electric utilities in a panel of ninety-one countries during the period from 1985 to 2018. I achieve two primary goals. First, I extend the Henisz et al. (2004) econometric study on electricity infrastructure deployment by increasing the number of datasets using ML techniques to predict capacity requirements. Second, I enhance Henisz et al. (2004) by using a firm, Eskom, rather than performing the study at country level, thus going beyond the econometric model of Henisz et al. (2004) in using an ML framework.

Henisz et al (2004) consider two factors that jointly influence the rate of infrastructure deployment:
(1) the extent to which the consumer base consists of industrial consumers, which are capable of exerting discipline on political actors whose competing incentives are to construct economically inefficient white elephants to satisfy the demands of concentrated geographic interests, labour unions and construction firms; and
(2) veto points (VP) in formal policymaking structures that constrain political actors, thereby reducing these actors' sensitivity to interest group demands.

The paper is structured as follows

Public Choice Literature

For the purpose of the discussion on public choice literature and its relevance to Eskom, I have split the public choice literature into two groups: the inefficiency of SOEs, and interest groups and veto points. Interest groups and veto points cut across various theoretical schools, as is discussed below.

Inefficiency of SOEs

First, Dyck (2001), Megginson & Netter (2000) and Vining & Boardman (1989) have emphasised the problem of inefficient resource allocation by considering both underinvestment and overinvestment in infrastructure by SOEs. On the one hand, public sector enterprises are viewed as being unable to raise sufficient funds to expand or even maintain the existing asset base. On the other hand, political actors seeking to curry favour with their constituents are argued to employ public sector enterprises, such as Eskom, to build white elephants - infrastructure projects of dubious economic necessity, but the construction of which benefits politically salient constituencies in the form of new jobs and associated multiplier effects in targeted sub-economies (Bertero & Rondi, 2000; Garrett & Lange, 1995; Karp & Perloff, 1989; and Shleifer & Vishny, 1994). For example, they would prioritise the interests of certain stakeholders (such as trade unions) at the expense of firm's efficiency. The proposed solution to both underinvestment and overinvestment in infrastructure is similar in spirit to that urged in the case of private sector investment distortions: to create institutional arrangements that reduce the extent to which political actors are able to impose their preferences on SOE investment patterns. However, Steyn (2006) shows that, in the case of Eskom, it was not the political actors that imposed preferences on SOE investment patterns but the managerial incentives and uncertainty of the political economic environment that led to inefficient overinvestment in capacity.

Second, the Agency Theory (Arrow, 1970; Ross, 1973) points to the separation of ownership and control as the main source of the relative poorer performance of public firms. Thus, it is argued that managers in privately-owned enterprises face stronger incentives to drive out waste and maximise internal efficiency. Conversely, the owners of public enterprises have a weaker ability to monitor the behaviour of managers. Managers in private companies are more disciplined by a number of external control mechanisms (e.g. the market for managers). There is some truth in that argument when institutions are insular and management has power to make investment decisions in SOEs outside the political process. One of the reasons for the appointment of the De Villiers Commission by PW Botha was to discipline Eskom, which was perceived to be too powerful, independent and unaccountable to political actors.

The **third** stream of thought is suggested by the property rights theorists (Barzel, 1997; Demsetz, 1974; and Alchian, 1965), who argue that managers in public firms do not suffer the economic consequences of their decisions, which reduces their incentives to reduce economic waste and maximise profitability. By contrast, the threat of bankruptcy and takeover prevent the managers of private firms from seeking only their own advantage. In this respect, the presence of soft budget constraints prevents public enterprises from bankruptcy, since any possible gap between income and expenditure is balanced by the government (Kornai, 1980).

Public choice theorists recommend the transfer of assets from the state to the private sector as a panacea for removing inefficiencies inherent to public ownership, which arguably reduces distortions and improves incentives. However, Christie's work (1978) has shown that in South

Africa, states and markets have mutually cooperated and coexisted in capitalist development. Although the two paradigms of state and private sector (markets) of the governance of the energy sectors can be complementary, there is consensus among policymakers that access to electricity is an essential ingredient for industrial development, which is considered a fundamental driver of economic growth.

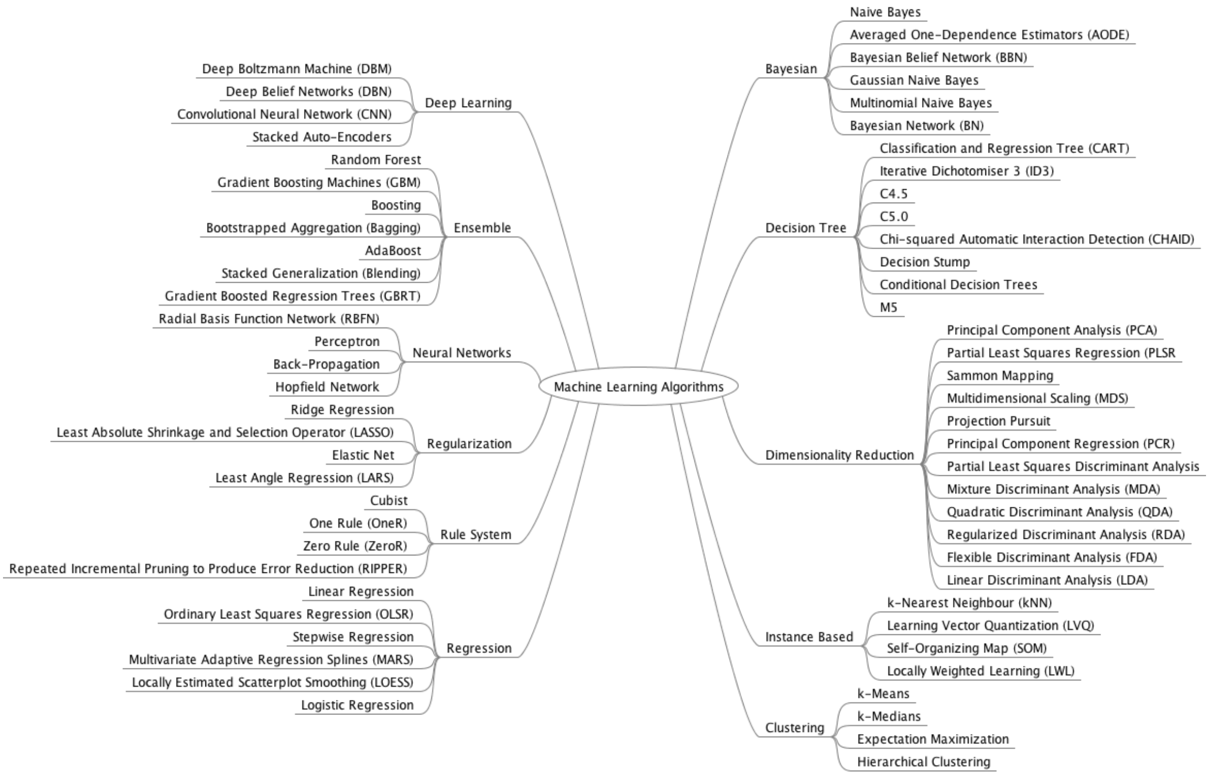
However, despite the theoretical arguments on the inefficiencies of SOEs, the empirical record of privatisation is mixed. The studies of the property rights theorists focus on the analysis of financial and operating indicators, typically profitability, labour productivity, sales, investment, debt, employment and dividends. The subsequent widely cited studies by D'Souza & Megginson (2002), La Porta & Lopez-de-Salines (1999) and Dewenter and Malatesta (1997) follow the same approach and generally find evidence that privatisation is associated with gains in firms' profitability and labour productivity, while delivering mixed results for other measures of performance.

Other studies provide evidence that profitability increases before privatisation (e.g. Alexandre & Charreaux, 2004; Dewenter and Malatesta, 2001), suggesting that governments can effectively restructure companies before selling them. In any case, most of these papers do not sufficiently control for the influence of other factors and economic changes that may occur in parallel and interact with the privatisation process (e.g. Boubakri, Cosset, Fischer, & Guedhami, 2005; D'Souza, Megginson, & Nash, 2005; Wallsten, 2001)). Megginson and Netter (2001) provide a survey covering 90 comparative studies that review performance changes of privatised companies, and Djankov & Murrell (2002) provide a quantitative investigation of firm restructuring in transition economies. A few other studies examine corporate performance after privatisation, using samples that also include some non-utility firms, and corroborate the superiority of privately-owned enterprises (POE) in terms of operating efficiency and profitability (Megginson et al., 1994; Boubakri and Cosset, 1998; D'Souza and Megginson, 1999; Dewenter and Malatesta, 2001).

The conventional argument that state-owned enterprises (SOEs) are inefficient compared with privately-owned ones (POEs) is sustained by the very well-established economic literature that we have discussed above. A large part of the empirical studies on the comparative performance of SOEs focuses on utilities, most often in non-competitive markets, where firms have a natural or special monopoly (typically, electricity and water utilities), or where there is a regulated duopoly (often this has been the case of airlines and railroads), or where output cannot be priced by competitive forces (e.g. health-related services). Notwithstanding a few exceptions, most available studies show that SOEs are less efficient than POEs, efficiency commonly being measured by short-term performance indicators such as returns on assets, returns on sales, and net income (see Vining and Boardman, 1989).

The public choice theory of free flow of goods that constitutes a laissez-faire economy on an infrastructural base that organises both market and society, has been challenged by scholars who draw on science and technology studies to trace out the material operation of technologies and the ways in which this materiality has consequences for political processes (Bennett, 2010; Callon, 1998; Latour, 2007, Mitchell 2011). Infrastructures, in this work, are interesting because they reveal forms of political rationality that underlie technological projects and which give rise to an “apparatus of governmentality” (Foucault 2010, p. 70).

Figure 1.1 Machine Learning and Simulink



Source: MATLAB R2018b

First, I identify the ML approach that best predicts electricity capacity deployment. Then, using the identified technique, I identify the variables that form the pattern that best predicts electricity capacity deployment. This is important because, at least theoretically, there may be reason to believe that many of the correlates of electricity capacity deployment have complex non-linear impacts on capacity deployment because strategic complementarities between variables can lead to multiple possible equilibria.

ML techniques identify tipping points in the range of a particular variable that may place a country in a lower or higher capacity deployment. Moreover, ML can generate partial dependence plots. These graphs can illustrate how variables will help us better understand causal patterns explaining growth. Moreover, I offer a better understanding of how electricity capacity deployment can be predicted, which will be of particular help to policy makers as they design policies for electricity infrastructure deployment.

The target variables in my analysis are provided in the four datasets that I integrate to test the hypothesis that a higher fraction of industrial customers provides political actors with stronger incentives for discipline, reducing the deployment of white elephants and thus the electricity infrastructure growth rate. Veto points reduce political actors' sensitivity to interest group demands in general and thus moderate the relationship between industrial interest group pressure and the rate of infrastructure deployment.

In contrast to Henisz et al. (2004) the search for variables is rather exhaustive, with robustly significant covariates. Thus, I begin with a long list of inputs from the World Development Indicators, British Petroleum (BP) World Energy Statistics, Statistics South Africa and POLCON from the Henisz et al. (2004) dataset. The POLCON dataset estimates the feasibility of policy change (the extent to which a change in the preferences of any one actor may lead to a change in government policy). It uses the following the following methodology.

First, extracting data from political science databases, it identifies the number of independent branches of government (executive, lower and upper legislative chambers) with veto power over policy change in [234] countries in every year [that they existed] from 1800 to [2001][1]. The preferences of each of these branches and the status quo policy are then assumed to be independently and identically drawn from a uniform, unidimensional policy space. This assumption allows for the derivation of a quantitative measure of institutional hazards using a simple spatial model of political interaction. This initial measure is then modified to take into account the extent of alignment across branches of government using data on the party composition of the executive and legislative branches. Such alignment increases the feasibility of policy change.

The measure is then further modified to capture the extent of preference heterogeneity within each legislative branch which increases (decreases) the decision costs of overturning policy for aligned (opposed) executive branches. Finally, the POLCON dataset includes information on the identity of the head of state and government (President Cyril Ramaphosa), the partisan affiliation of the head of state and government (African National Congress) and the seat distribution by party in the upper and lower legislative chambers (Democratic Alliance, Economic Freedom Fighters, Inkatha Freedom Party, United Democratic Movement, National

[1] For a list of countries, see appendix 1.

Freedom Party, African Christian Democratic Party, Pan Africanist Congress of Azania, Azanian People's Organisation, Freedom Front Plus, Congress of the People, African Independent Congress, Agang South Africa, African Independent Congress and African People's Convention. (Henisz W. J., 2002). For example, in the case of South Africa, the POLCON dataset starts when the head of state was General Botha and continues until 2017, under President Zuma (See Appendix 1) .

ID = An auto-numbered index or key
CNTS_COUNTRY = Country name for matching to the cross-national time series dataset: http://www.databanks.sitehosting.net/
POLITY_COUNTRY = Country name for matching to the Polity datasets: http://www.bsos.umd.edu/cidcm/inscr/polity/index.htm
ICRG COUNTRY = Country name for matching to the International Country Risk Guide datasets: http://www.icrgonline.com/
CTRYNM = Country code for matching to any World Bank datasets
CNTS_CODE = Country code for matching to the cross-national time series dataset: http://www.databanks.sitehosting.net/
YEAR = Year (data as of January 1 of that year): http://www.bsos.umd.edu/cidcm/inscr/polity/index.htm

Next, I include lagged values of several variables to account for industrial representation from the World Development Index dataset. The *World Development Indicators* (WDI) is the World Bank's premier compilation of international statistics on global development. Drawing from officially recognised sources and including national, regional, and global estimates, the WDI provides access to almost 1,600 indicators for 217 economies, with some time series extending back more than 50 years. The database helps users to find information related to all aspects of development, both current and historical. The WDI are organised according to six thematic areas: *poverty and inequality*, *people*, *environment*, *economy*, *states and markets*, and *global links*. Each thematic page provides an overview of the type of data available, a list of featured indicators, and information about widely-used methodologies and current data challenges. I chose the economy thematic area.

The economy dataset provides a window into the global economy and the economic activity of the more than 200 countries and territories that produce, trade, and consume the world's output.

The indicators in the WDI database monitor changes in the size and structure of the global economy and their effects on national economies. Indicators include measures of macroeconomic performance (GDP, consumption, investment, international trade, balance of payments, central government budgets, prices, and money supply). Broader measures of income and savings, adjusted for pollution, depreciation, and depletion of resources, are also available. I then include several lagged values pertaining to energy production, consumptions, sales and other variables from the *BP World Statistics Review* 2018 dataset.

BP publishes an annual review of energy statistics. The 2018 statistics published in the review were taken from government sources and published data. The statistics are sourced from 91 countries, divided into various blocs, as shown in Appendix 2. Finally, I consider variables that are South African, relating to electricity capacity deployment from the Statistics South Africa dataset. The Stats SA dataset consists of electricity production, consumption, sales, and exports imports across all energy forms (oil (diesel), wind, hydro, solar and coal (thermal)).

Dependent variable

The dependent variable in the conceptual hypotheses that I advance is a country's rate of white elephant deployment. Empirically, it is not possible in a wide panel of countries to separate the deployment of white elephant capacity from that of economically justifiable deployment without subjectively assessing extremely detailed data on investment costs and reserve ratios (Henisz & Zelner, 2004). However, we do not require such a measure to test our hypotheses. Rather, the marginal nature of these hypotheses, noted in their development above, permits us to use objective data on the annual growth rate of total SOE generating capacity for our estimated GLM. ML techniques will allow us to assess how political variables corresponding to those in our hypotheses increase or decrease the annual rate of infrastructure deployment when the economic determinants in our model are taken into account. The operative question is, when economic determinants are taken into account, do the political variables have explanatory power, and if so, what are the direction and magnitude of their influence? Statistical insignificance of the political variables would refute our hypotheses, as would coefficients with a sign or relative magnitude inconsistent with the hypotheses (Henisz & Zelner, 2004).

Independent variables

Industrial representation. We measure industrial representation as the one-year lagged ratio of industry value added as a percentage of GDP. Data used to construct this measure are reported by the *BP World Energy Statistics*.

Veto points. We measure the level of VP affecting political actors in terms of the structure of a country's formal political institutions and the extent of partisan heterogeneity within and among these institutions. I employ Henisz's Political Constraints Index (POLCON) in the identification of the number of independent branches of government (executive, lower and upper legislative chambers, judiciary and sub-federal institutions) with veto power over policy change in each country (See Appendix 2). Countries with the greatest level of VP in the formal policymaking apparatus are those federal states with strong independent judiciaries and either presidential systems or proportional representation electoral rules that tend to yield coalition governments, such as the United States, Germany and Switzerland (Henisz & Zelner, 2004).

Political constraints decrease as the number of veto players declines or as their preferences become more homogeneous, as is the case in moving to a mixed parliamentary-presidential system, typified by France or Brazil; to heavily fractionalised parliamentary systems like those of Belgium, Israel and the Netherlands; to Westminster parliamentary systems with winner-take-all districts, such as the United Kingdom's. Non-democratic countries and those with transitional political regimes have the lowest levels of political constraints because the formal institutional structures in these states provide tremendous discretion to policymakers (Henisz & Zelner, 2004).

Existing capacity level. The one-year lagged value of the existing level of capacity per capita (CAPACITYPC) reflects the effect of two influences on the rate of deployment. First, CAPACITYPC measures the economic demand for replacement stock, and should therefore be negatively correlated with the rate of new infrastructure deployment. Second, where existing capacity is low, political actors seeking to build white elephants are more easily able to assemble broader political support for the deployment of new infrastructure. Under a higher level of existing capacity, on the other hand, these political actors find it more difficult to assemble broad support because they can no longer as easily make the case for new capacity by appealing to the common interest of their colleagues.

As a result, we expect the negative influence of CAPACITYPC to be conditional on that of the political variables of central interest. As IR rises, the incentives that political actors face to exert discipline strengthen, implying that political actors will use a given increase in CAPACITYPC to argue for a greater amount of discipline, leading to a lower rate of deployment of new capacity. That is, the negative marginal effect of CAPACITYPC on the deployment rate should decline as IR rises.

Similarly, the magnitude of the negative marginal effect of CAPACITYPC on the rate of deployment should also depend on the level of VP that political actors face. A political actor's arguments for discipline in the case of a given increase in CAPACITYPC are less likely to result in approval as the number and breadth of interests of the veto players among which agreement must occur grows. Thus, the magnitude of the negative effect of CAPACITYPC on the deployment rate should decline as VP increases. In have used the natural logarithm of CAPACITYPC because the distribution of the variable's raw levels is skewed to the left.

Demand. It is critical to control for the economic demand for new infrastructure. This demand derives from the expected future demand for electricity. However, actual forecasts of expected demand are unavailable for most countries and time periods, and in any case pose the issue of endogeneity in a model for which the dependent variable is the rate of capacity deployment. I use consumption, measured as the prior year's end-user electricity consumption measured in kilowatt hours per capita (DEMANDPC), to proxy for the (unobservable) demand for infrastructure. Recent consumption is clearly exogenous to infrastructure deployment choices, and political actors observe this measure when making deployment choices (Henisz & Zelner, 2004).

Financial constraints. I use public debt levels as a percentage of GDP since it could increase the cost of infrastructure deployment and thereby reduce the rate of deployment, ceteris paribus. As this variable declines, reflecting a decline in the country's cost of capital, we expect to observe a higher rate of capacity deployment.

Availability of foreign supplies. Some governments can buy electricity from abroad rather than generate it domestically. Our specification therefore includes the lagged ratio of imported electricity to total electricity consumed (IMPORTRAT). An alternative source of supply should negatively affect the generating capacity deployment level. In our specification, we use the natural logarithm of IMPORTRAT because the distribution of raw levels of the variable is skewed to the left.

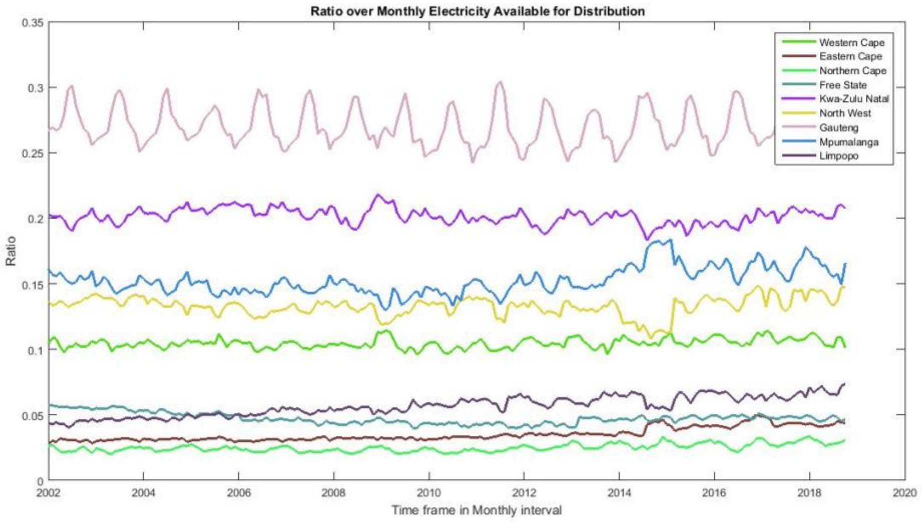
Composition of domestic supplies. The composition of a country or region's generating capacity by technology type (nuclear, coal, oil, gas, hydroelectric and others) is influenced by the magnitude and frequency of daily and seasonal fluctuations in demand. In general, larger baseload generating units, with low unit costs and high start-up costs, more efficiently serve relatively stable demand; and smaller peaking units, which are more expensive to run but can also be taken online and offline more easily, more efficiently serve uneven demand (Henisz & Zelner, 2004). Although differences in capacity composition should not affect total capacity in a steady-state equilibrium, changes in composition may affect the total level of capacity during transition periods. A strong shift in daily or seasonal demand patterns, for example, might render it efficient to increase baseload capacity, while maintaining existing peaking capacity, but keeping it idle most of the time.

Model development

I have identified a linear model for our ML approach to predict electricity capacity deployment. I will start with a simple model using Stats SA data on the South Africa SOE, Eskom. This is different from the approach of Henisz et al. (2015) whose unit of analysis is a country, since our first model build uses the firm data of Eskom. I want to learn how the machine behaves on a monotype dataset. Generally, successful modelling requires a thorough analysis of data. In the data, a number of covariates (variables) are examined prior to model fitting to familiarise yourself with the data in order to enable grounded data-backed arguments. Prior to fitting a generalised linear model (GLM) on the Eskom data, several significant variables were first extracted and examined. The covariates include: CAPACITY (Electricity produced), IMPORTRAT (Electricity produced outside South Africa), INTERNAL_CONSUMPTION (Electricity consumed in power stations and auxilliary systems), EXPORTRAT (Electricity available for distribution outside South Africa) and DEMAND (Electricity available for distribution in South Africa). For the current objective, we do not perform the scaling of variables because the Eskom data is monotype.

However, at a country level there are a variety of contributors and the variety of parameters which causes data to be dispersed, as will be shown in Model B below. To enable a comparison, I scale the data by calculating ratios and presenting all variables as decisive ratios so as to have mathematically valid arguments. I examine the demand parameter by examining the electricity consumed by each of the nine provinces of South Africa as a percentage of the total electricity that was available for distribution at the specified time. Therefore the equation is derived as amount of electricity allocated to province x at time t divided by electricity available for distribution at the time i.e. $Ratio_t = \frac{ELK_t{}^x}{ELK_t{}^{AFD}}$ where x is any element of the set {Gauteng, Western Cape, Eastern Cape Northern Cape, Limpopo, North West, Mpumalanga, Free State, KwaZulu-Natal}. The results are indicated graphically below in Figure 1.2 as in the below:

Figure 1.2 : The ratio of the electricity consumed in each province out of the total energy available at the time.



Source: Stats SA

Upper quantile:

Gauteng is the largest consumer of electrical energy, accounting for between 25 and 30% of the total available energy throughout time. It is followed by KwaZulu-Natal, which accounts for about 20%.

Middle quantile:

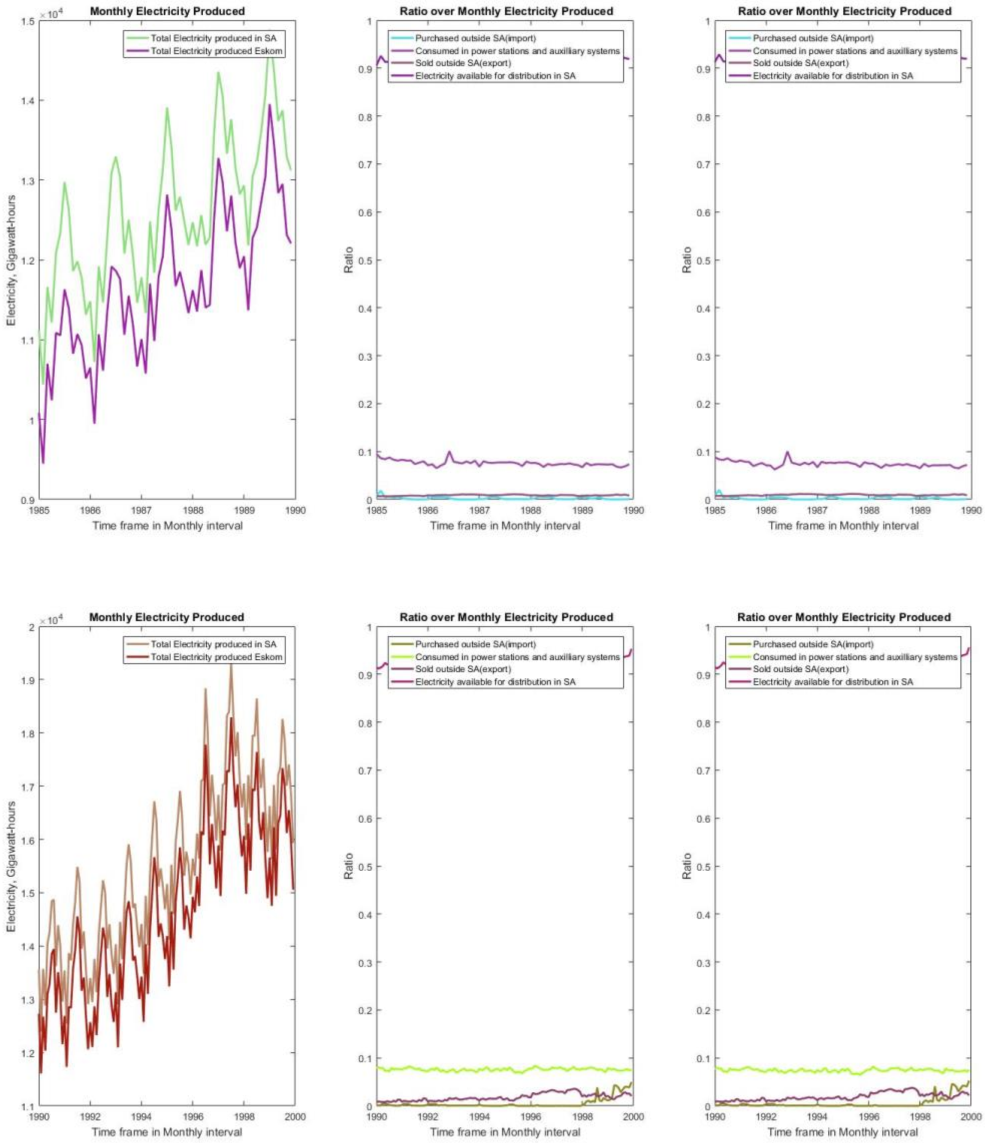
Mpumalanga accounts for about 15%; North West accounts for about 13%, and the Western Cape accounts for 10%.

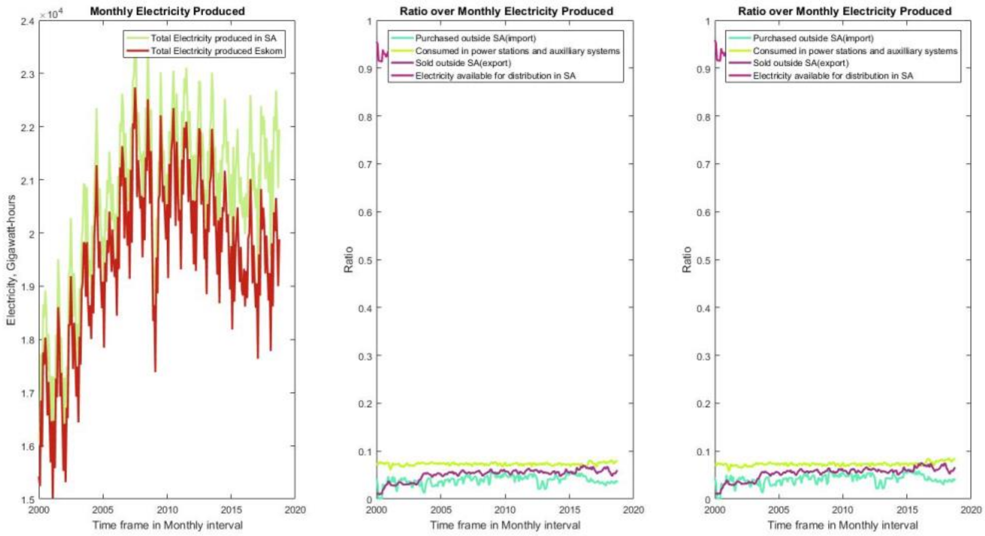
Lower quantile:

Limpopo demand began below 5%, and increased steadily; contrarily, Free State demand began above 5% and decreased steadily. The Eastern Cape and the Northern Cape have highly correlated demand, which began at around 3% and increases steadily.

However, at a country level, a variety of contributors and the parameters causes data to be dispersed. To enable us compare variables, I scale the data by calculating ratios and presenting all variables as decisive ratios to enable us to have mathematically valid arguments. I then graphically plot them, as shown in Figure 1.3 below.

Figure 1.3 CAPACITY, IMPORTRAT, INTERNAL CONSUMPTION, EXPORTRAT, and DEMAND 1985 -2017





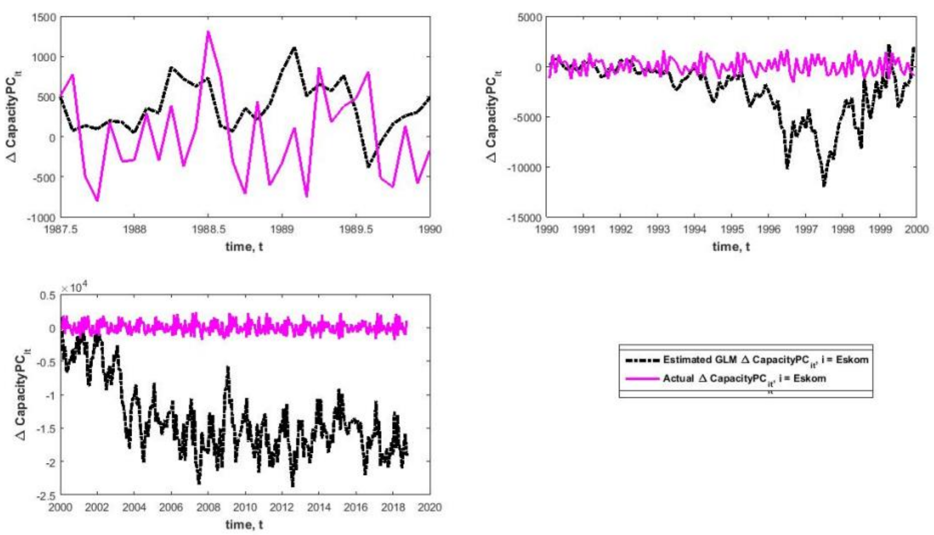
Subsequently, a linear model is trained and tested over the data from 1985 to 2018. For the training part, the data from 1985 to 1989 was subdivided into two parts. The lefthand part was applied to train the model and the righthand part was applied to test the model performance. The reason the training was done with the 1985 to 1989 data is its reliability and consistency.

Table 1.1: Learning Sample and Test Sample

		Learning Sample					Test Sample			
Source	Obs	Mean	Std	Min	Max	Obs	Mean	Std	Min	Max
LnCAPACITYPC	29	9,393379765	0,059864676	9,253304071	9,49499286	31	9,487469	0,055226168	9,406893511	9,607974901
WASTE	29	33,03448276	39,84925412	0	190	31	23,06452	24,28159177	0	76
IMPORTRAT	29	944,1034483	106,4117155	777	1319	31	967,3548	62,28298237	827	1082
EXPORTTRAT	29	102,3793103	17,63603007	69	128	31	116,6774	14,1005605	90	139
ELECTRIC_FOR_DIST_SA	29	11015,82759	624,2291285	9662	12130	31	12151,94	671,6074714	11116	13678

In the top left graph of Figure 1.3 below, I plot the actual lagged capacity change of Eskom and the estimated GLM lagged capacity change of Eskom. The results as predicted by the machine and the actual change in capacity have been plotted on the same scale to enable comparison. Clearly, the machine produces a satisfactory performance, as indicated in the top left of the graphs of Figure 1.4. Both the actual and the machine estimated data fall within the same bounds.

Figure 1.4: The machine estimated change in capacity and the reported change in capacity as per the Eskom annual financial statements (1985-2018).



Source: Stats SA

The consistency is carried through until around mid 1993 to 1994. Then the inconsistency begins (false reports). Around 1999, the reporting is estimated to be accurate again. Strictly speaking, as from the beginning of 2002 onwards, Eskom has been under-performing up to date. However, Eskom has recorded prominent numbers in its reporting, contrary to what ML implies. It shows Eskom's numbers to be less reliable as from the beginning of 1994 and strongly so subsequent to 2002. This finding is supported by experiences in the load shedding in the 2007/2008 period and the liquidity and sustainability issues that have beset the organisation. The Generalised Linear Model (GLM) as presented above appears to be forward looking. I have fitted the model according to Henisz et al.(2004), except the on the scaling part.

The scaling was left out because the data is monotype. However, for country data, scaling plays a critical role in aligning it, as it will be discussed below. The trained and tested model is provided mathematically as:

$$\begin{aligned}\Delta \mathrm{CAPACITYPC}_{i,t} &\\ &= \alpha_0 + \alpha_1 \ln\mathrm{CAPACITYPC}_{i,t-1} + \alpha_2 \mathrm{IMPORTRAT}_{i,t-1} \\ &+ \alpha_3 \mathrm{INTERNAL\_CONSUMPTION}_{i,t-1} + \alpha_4 \mathrm{EXPORTRAT}_{i,t-1} \\ &+ \alpha_5 \mathrm{DEMAND}_{i,t-1}\end{aligned}$$

Where $\Delta CAPACITYPC_{i,t}$ is the one-year lagged value of the existing level of capacity per capita (CAPACITYPC), which measures the economic demand for replacement stock.

α_0 = Coefficient (y intercept)
$\alpha_1 \ln\mathrm{CAPACITYPC}_{i,t-1}$ = The lagged ratio of the prior year's level of capacity per capita.
$\alpha_2 \mathrm{IMPORTRAT}_{i,t-1}$ = South is part of the South African Power Pool where electricity produced is traded. Cahora Basa is a source of electricity for South African consumption. The specification therefore includes the lagged ratio of imported electricity to total electricity consumed.
$\alpha_3 \mathrm{INTERNAL\_CONSUMPTION}_{i,t-1}$ = The lagged ratio of domestic consumption of electricity.
$\alpha_4 \mathrm{EXPORTRAT}_{i,t-1}$ = The lagged ratio of exported electricity to total electricity consumed.
$\alpha_5 \mathrm{DEMAND}_{i,t-1}$ = Consumption of electricity, measured as the prior year's end-user electricity consumption measured in kilowatt-hours per capita (DEMANDPC), to proxy for the (unobservable) demand for infrastructure.

Table 1.2 below reports the estimation results for the core specification described above and several variants that we use to assess the results' robustness. Column one reports results of the coefficients from α_0 to α_{05}. The coefficient estimate for each variable or interaction term is individually significant at a p-value of 0,05 or less. However, the individual point estimates of the coefficients $\mathrm{INTERNAL\_CONSUMPTION}_{i,t-1}$, $\mathrm{EXPORTRAT}_{i,t-1}$ and $\mathrm{DEMAND}_{i,t-1}$, the independent variables of primary interest, do not have a meaningful interpretation as a result of the interaction terms in the model. Rather, proper assessment of the effects of

$\text{INTERNAL\_CONSUMPTION}_{i,t-1}$, $\text{EXPORTRAT}_{i,t-1}$ and $\text{DEMAND}_{i,t-1}$ on the capacity growth rate (ΔCAPACITYPC) depends on the respective estimators.

Table 1.2 Coefficients of the GLM

	Estimate	SE	tStat	pValue
Intercept	-7,9212e+5	5,2804e5	-1,5001	0,1472
x_1	9,3472e+4	6,2924e+4	1,4855	0,1510
x_2	13,5748	7,8719	1,7245	0,0980
x_3	-8,0662	4,9794	-1,6199	0,1189
x_4	-15,8344	9,9121	-1,5975	0,1238
x_5	-6,9936	5,3090	-1,3173	0,2007

Rsquared

Ordinary	Adjusted
0,3817	0,2472

MSE	3,6964e+5
RMSE	607,9774
LogLikelihood	-221,4046
DFE	23
SSE	8,5016e+6
SST	1,3749e+7
SSR	5,2475e+6

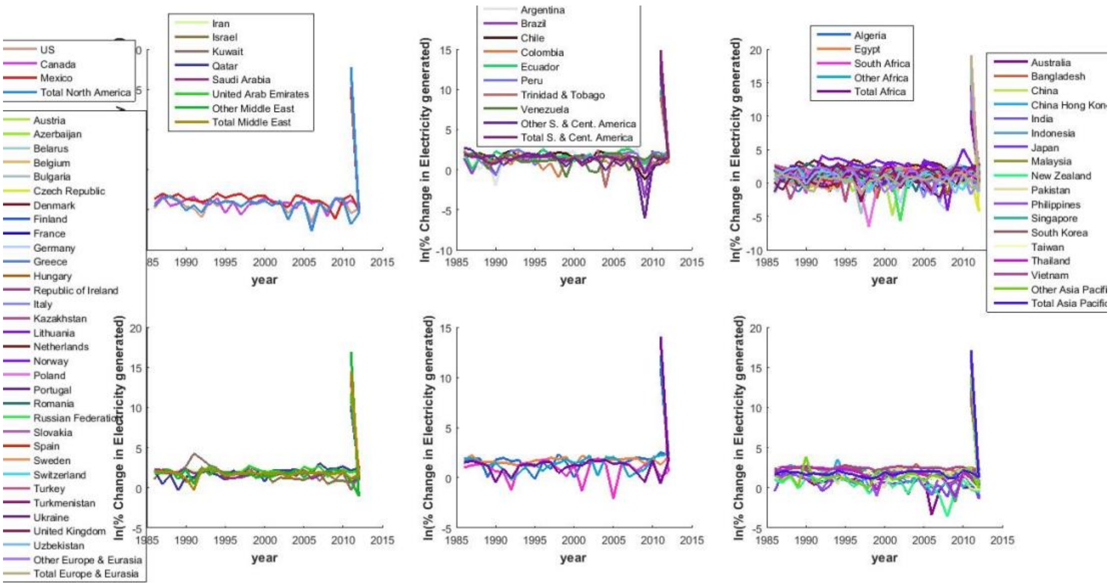
The results show an improved R squared measure of 0,2472 over that of the Henisz et al. (2004) which recorded an R squared quantity of about 0,22 based on the country as unit of analysis. However, since I have not yet introduced the political economy variables of Henisz, any arrival to conclusions about the hypotheses that I have presented would be premature. However, the model is substantially strong and shows that the ML approach has more variability in explaining the deployment of electricity capacity and is robust when compared with the literature methodologies, which rely solely on macroeconomic indicators. Also noticeable are the Mean Squared Errors (MSE) above, which are significantly contained compared with the MSE of the Henisz et al. (2004). Such an observation suggests that the monotype data, when injected in a linear model, better explains the behaviour of Eskom in terms of change in capacity when compared with the use of macro-variables as proxies as per the approaches in the economic literature (Inglesi-Lotz & Pouris, 2016, 2011; Inglesi, 2010a; 2010b, Amusa, Amusa, & Mabugu, 2009). The coefficient covariance, variable info and model criterion can be found in Appendix 3.

Testing the hypothesis - Analysis of electricity for South Africa versus the rest of the world

Having shown the robustness of ML techniques as shown by the GLM model above, I now use the South African electricity data as reported in the WDI of the World Bank, the *BP World Energy Statistics*, and the POLCON dataset of Henisz et al. (2004) to test the hypothesis and to determine whether Henisz et al. holds when using ML techniques. The country (South Africa) is the unit of analysis. I start by calculating the change in total electrical energy produced per country as the percentage of the one-year lagged total electrical energy produced. The resultant distribution is skewed to the left and as a result there are no clearly observable insights. I then multiply by 100 and take the natural log to visualise results. Surprisingly, the results follow similar structures, irrespective of the country analysed. The results are shown in Figure 1.5.

Figure 1.5 shows all the results in one footage. From the graphs, we observe how demand for energy through different countries is governed by the same statistical rule, with specified states being correlated to others. This observation indicates that the parameters of a specified country may be estimated using the data of another country. More importantly, the main technique that allows such comparison and conclusion to occur is taking ratios of all inputs and outputs when doing an analysis for a specified country. This is because, in absolute terms, the data from one country cannot be superimposed on another country. Also, it is difficult to compare and derive mathematically sound results. However, when each covariate is scaled and ratios of all covariates are being considered instead of absolute numbers, comparison is possible and robust results that are mathematically valid could be derived.

Figure 1.5: The natural log of percentage change for countries in all the countries in the BP World Energy Statistics



To create the ratios, I adopted the methodology of Henisz et al. (2004) for each parameter considered. Otherwise I scale by the mean of covariate where applicable. The most important conclusion is that the demand for electrical energy is governed by the same rule and as such any internal improvement in the energy supply of one country may be improved by studying and implementing the approaches of other, better performing countries. As in Model A, where I used Eskom and Stats SA data, using the learning sample and the test sample as shown in Figure 3.6. I repeat the same approach with the comprehensive datasets discussed above.

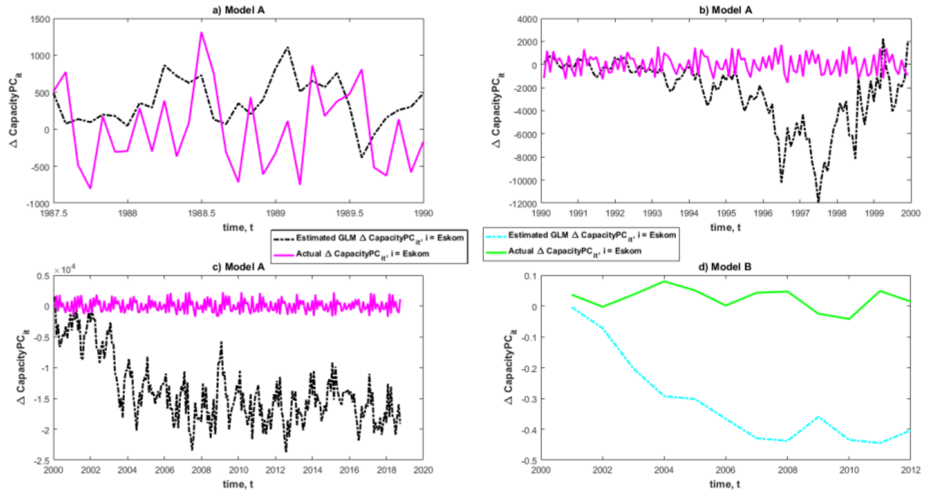


Figure 1.6: Model suggested versus reported change in capacity

In Figure 1.6, I show the original Model A of Stats SA and Eskom data on the left top and bottom and on the right the Model B of the comprehensive data sets. Both Model A and Model B are robust in predicting the decline in electricity deployment as shown by the Estimate GLM $\Delta CAPACITYPC_{t-1}$ in both models. The models show that there was a mismatch between capacity available and demanded and therefore Eskom, in Model A, and South Africa, in Model B, should have deployed new electricity capacity.

Table 1.3: Learning and test sample in machine learning

	Learning Sample					Test Sample				
Source	Obs	Mean	Std	Min	Max	Obs	Mean	Std	Min	Max
LnCAPACITYPC	15	-0,149079184	0,124634149	-0,352145602	0,018700027	12	0,172917	0,07378943	0,015991	0,242380768
POLSA	15	0,277307987	0,165001454	0	0,46273432	12	0,431053	0,02348289	0,412194	0,46273432
IR	15	0,233847041	0,017315392	0,209348391	0,261773894	12	0,319795	0,01395497	0,298898	0,344344934
POLSA_x_IR	15	1310009,553	97001,66251	1172766,757	1466457,356	12	0,13767	0,0061008	0,127903	0,146837448
POLSA_x_LnCAPACITYPC	15	-984083,317	721291,4904	-2129389,43	96561,38912	12	0,073207	0,02977485	0,007399	0,101078505
IR_x_LnCAPACITYPC	15	-0,039278953	0,027224704	-0,07957611	0,004178015	12	0,055727	0,02436577	0,004887	0,077970851

POLSA_x_IR_x_LnCAPACITYPC	15	-185455,8521	149134,1912	-412984,6558	27422,79252	12	0,023572	0,00981698	0,002261	0,032139131
LnDEMAND	15	8,188136998	0,143963712	7,926937374	8,337363758	12	8,36281	0,06622072	8,275685	8,464761848
INVEST	15	521888818,3	297903141,4	41480518,22	1021379053	12	6,34E+08	341906618	18876127	1049153909
IMPORTRAT	15	88,6	11,60541745	69	106	12	966,3333	56,5224546	869	1072

LnCAPACITYPC= Natural log of percentage change in capacity

POLSA = Political Constraints Index for South Africa

IR = Industry value added as a percentage of GDP

POLSA x IR = Political Constraints Index for South Africa multiplied by industry value added as a percentage of GDP

POLSA x LnCAPACITYPC = Political Constraints Index for South Africa multiplied by natural log of percentage change in capacity

POLSA x IR xLnCAPACITYPC = Political Constraints Index for South Africa multiplied by natural log of percentage change in capacity

LnDEMAND = Natural log of percentage change in demand for electricity

INVEST = Investment as a percentage of GDP

IMPORTRAT = Change in electricity produced outside South Africa

Table 1.4: Coefficients

		Coefficients		
Covariate	Estimate	SE	tStat	pValue
Intercept	0,060665362	0,400237186	0,151573526	0,886860708
x1	-0,384670308	0,785981185	-0,489414143	0,650173953
x2	-0,001290408	0,008953312	-0,144126317	0,892370514
x3	0	0		
x4	-3,1924E-08	2,36637E-08	-1,349068947	0,248624286
x5	7,22188E-08	1,43011E-07	0,504986045	0,640121061
x6	-5,451162618	3,288171666	-1,65780962	0,172695825
x7	9,76674E-07	5,83154E-07	1,674812514	0,169282223
x8	-0,002819713	0,047470335	-0,05939947	0,955483114
x9	2,31999E-12	2,44056E-12	0,950598372	0,395637477
x10	7,29865E-05	7,52026E-05	0,970530975	0,386739884

Error	Measures
RsquaredOrd.	0,996996237
RsquaredAdju.	0,991589463
AIC	-140,273123
MSE	4,01793E-06
RMSE	0,002004478

The results show an improved adjusted R squared measure of 0,991 over the Henisz et al. (2004) results, which recorded an R squared quantity of about 0,22 based on the country as unit of analysis. However, since I have not yet tested the political economy variables of Henisz, any arrival to conclusions about the hypotheses that I have presented would also be premature. However, the model is substantially strong and shows that the ML approach has more variability in explaining the deployment of electricity capacity and is robust when compared with the literature methodologies, which rely solely on macroeconomic indicators. The mathematical equation of the model is presented as:

$$
\begin{aligned}
\Delta \mathrm{CAPACITYPC}_{i,t} \\
&= \beta_0 + \beta_1 \mathrm{lnCAPACITYPC}_{i,t-1} + \beta_2 \mathrm{VP}_{i,t-1} + \beta_3 \mathrm{IR}_{i,t-1} \\
&+ \beta_4 \left(\mathrm{VP}_{i,t-1} \times \mathrm{IR}_{i,t-1}\right) + \beta_5 (\mathrm{VP}_{i,t-1} \times \mathrm{lnCAPACITYPC}_{i,t-1}) \\
&+ \beta_6 \left(\mathrm{IR}_{i,t-1} \times \mathrm{lnCAPACITYPC}_{i,t-1}\right) \\
&+ \beta_7 (\mathrm{VP}_{i,t-1} \times \mathrm{IR}_{i,t-1} \times \mathrm{lnCAPACITYPC}_{i,t-1}) + \beta_8 \mathrm{lnDEMAND}_{i,t-1} \\
&+ \beta_9 \mathrm{CAPCOST}_{i,t-1} + \beta_{10} \mathrm{lnIMPORTRAT}_{i,t-1}
\end{aligned}
$$

The mathematical equation

The mathematical representation that was applied to calculate the results of figure 1.6 is provided by :

$$y_{i,t} = \ln(\frac{e_{i,t} - e_{i,t-1}}{e_{i,t-1}} \times 100)$$

where i∈ $\{US, SA, Mexico, Canada, \ldots\}$. That is (i) represents a specified country.

The other part of the analysis discovers that South Africa's power utility is perfectly correlated to Eskom. Figures 1.7 to 1.9 substantiate the claim.

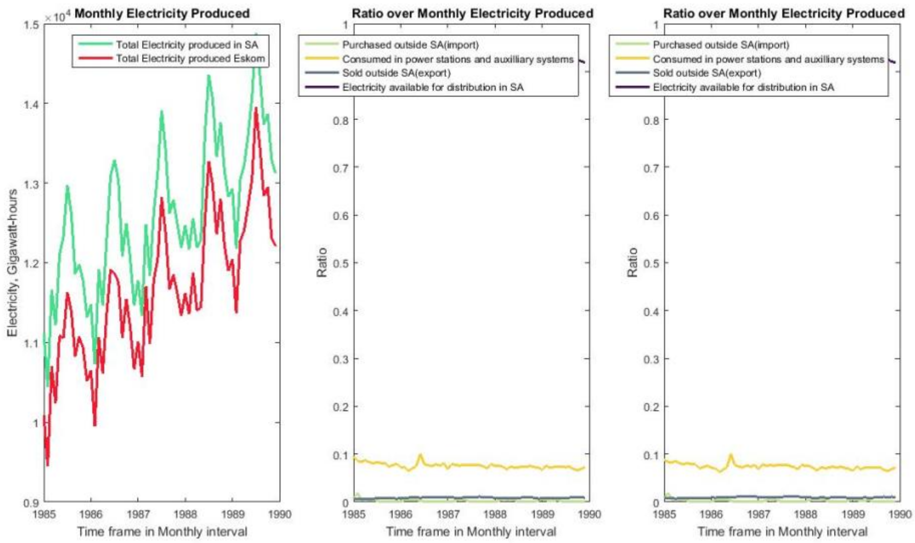


Figure 1.7: Eskom's total capacity and the ratio proportions of energy allocations.

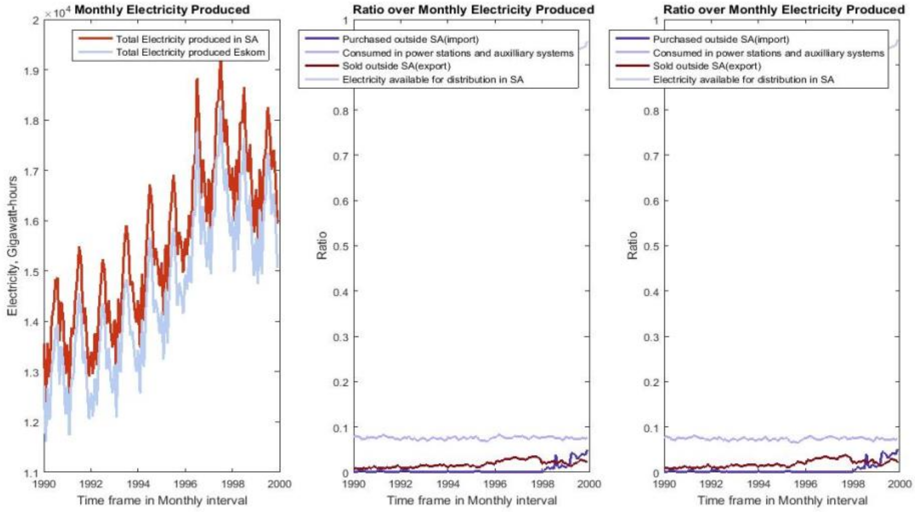


Figure 1.8 : Eskom's total capacity and the ratio proportions of energy allocations

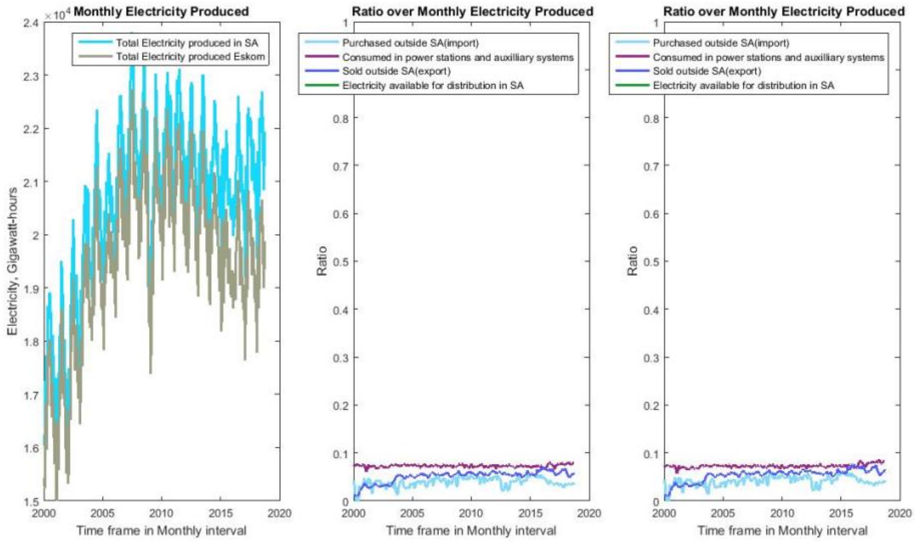


Figure 1.9: Eskom's total capacity and the ratio proportions of energy allocations

Discussion

At the beginning of this paper, we set out to test the validity of three hypotheses:

H1: The rate of white elephant deployment declines as the level of industrial representation in the consumer base rises, ceteris paribus.

H2: As the level of institutional VP in the policymaking process increases, the absolute magnitude of the negative relationship between industrial representation and the rate of white elephant deployment declines, ceteris paribus.

H3: An increase in the level of VP reduces the rate of white elephant deployment.

Table 1.4 reports the estimation results for the core specification described above and several variants that I have used to assess the results' robustness. Column one reports the covariate intercepts (variables) and column two contains the estimated coefficients of the covariate. Consider the first column. With the exception of certain time periods, the coefficient estimate for each variable or interaction term is individually significant at a p-value of 0,05 or less. However, the individual point estimates of the coefficients for IR and VP, the independent variables of primary interest, do not have a meaningful interpretation as a result of the interaction terms in the model. Rather, proper assessment of the effects of IR and VP (as well as CAPACITYPC) on the capacity growth rate (ΔCAPACITYPC) depends on the respective estimators and their standard error.

Hypothesis 1 states that the marginal effect of IR should be negative (regardless of the level of VP). Consistent with this hypothesis, the effect of IR on the rate of infrastructure deployment when VP (measured here by POLCON) takes any value from its sample minimum to sample maximum is negative (-3,1924E-08) and has a p-value of 0,25 or less at all reported values of VP except for the sample maximum.

Hypothesis 2 states that the marginal effect of IR should decline in absolute magnitude as VP increases. When VP is set to its sample mean minus one standard deviation (a low level of VP), a one standard deviation increase in IR (0,23) yields a predicted decline in the infrastructure deployment rate of 15 percentage points (the slope coefficient of -0,69 multiplied by the increase in IR of 0,23), equal to 164% of the absolute value of the infrastructure deployment rate's sample mean of 4,4 percentage points (or 58% of one standard deviation of the infrastructure deployment rate).

When VP rises to its sample mean value (0,39), the effect of an increase in IR of one standard deviation (0,16) declines in absolute magnitude to a predicted reduction of 5,0 percentage points (the slope coefficient of -0,31 multiplied by the increase in IR of 0,16), equal to 136% of the absolute value of the infrastructure deployment rate's sample mean (37,5% of one standard deviation). The negative marginal effect of pressure exerted by the lobby favouring discipline therefore declines in absolute magnitude as the level of VP rises. Moreover, when the level of VP reaches its sample maximum, this effect declines so much in absolute magnitude that it is statistically indistinguishable from zero ($p = 0,21$). These results are consistent with Hypothesis 2: as the level of VP imposed by formal institutional structures increases, the negative marginal influence of industrial consumers on infrastructure deployment declines.

Hypothesis 3 addresses the effect of VP when IR = 0. When IR takes its sample minimum value of 0,06, the marginal effect of VP is negative, with an estimated value of -0,17 and a p-value of 0,06. When IR takes the out-of-sample value of zero, the implied marginal effect of VP is -0,18, suggesting that a one standard deviation increase in VP (0,35) would generate a predicted decrease in the infrastructure deployment rate of 6,3 percentage points, equal to 155% of the absolute value of the deployment rate's sample mean (or 52% of one standard deviation). These results are consistent with Hypothesis 3: when political actors are predisposed to cater to the white elephant lobby's demands for increased deployment, the marginal effect of VP – which reduces political actors' sensitivity to interest group pressures— is a reduction in the deployment rate, ceteris paribus.

Although we have not proposed a formal hypothesis about the effect of VP on the rate of deployment at levels of industrial representation (IR) exceeding zero, the pattern of effects of VP at different values of IR is nonetheless informative. First consider the effect of an increase in VP when IR takes its sample mean value. This effect is statistically indistinguishable from zero (the estimate of 0,01 has a p-value 0,84). Thus, where there exists some intermediate level of industrial representation (IR) and the level of pressure to build white elephants is held constant, an increase in VP has no observable effect on the rate of infrastructure deployment. The lack of an observable effect is consistent with the proposition that such an increase reduces the sensitivity of political actors to both the pro-white elephant and pro-discipline lobbies.

When industrial representation (IR) rises to its sample maximum value of 0,92, the estimated marginal effect of VP is 0,19 with a p-value of 0,01, implying that a one standard deviation increase in VP would now increase the predicted infrastructure deployment rate by 6,7 percentage points. The marginal effect of VP – which reduces political actors' sensitivity to interest group pressures – becomes positive as the strength of the pro-discipline lobby decreases, and thus the predisposition of political actors to cater to this lobby's demands for reduced deployment, grows.[60]

(CAPACITYPC), however, it is necessary to calculate BCAPACITYPC (analogous to BIR and BVP) because CAPACITYPC is the third variable included in the interaction terms. When VP and IR are permitted to vary within one standard deviation around their sample mean, the total effect of existing capacity level ranges from -0,198 to -0,217, with a p-value of 0,008 or less. This effect persists in both sign and statistical significance over almost all feasible combinations of values of the three variables,[61] and thus supports the conjecture advanced above that the deployment rate is inversely related to the existing capacity level.

Demand. The coefficient estimate for DEMANDPC is significant and positively signed, indicating that countries with higher levels of electricity consumption build more capacity. It is informative to compare the magnitude of the estimated coefficients of these economic control variables to those of the political variables of theoretical interest. DEMANDPC is a central economic driver of infrastructure deployment; indeed, as noted above, several previous cross-national studies employ a single measure of demand as their sole independent variable.
In our specification, an increase in DEMANDPC of one standard deviation (1,69), increases the predicted rate of capacity deployment by 24 percentage points, or two standard deviations of the dependent variable. However, when industrial representation (IR) rises to one standard deviation above its mean (the pro-discipline lobby is relatively strong) and the level of VP declines to one standard deviation below its mean (political actors are relatively sensitive to interest group pressure), the predicted negative marginal effect of industrial representation on the rate of capacity deployment is -29 percentage points, which more than offsets the positive marginal effect of DEMANDPC to reduce the predicted rate of capacity deployment by five percentage points. In contrast, when IR declines to one standard deviation below its mean (a relatively weak pro-discipline lobby), the negative effect of industrial representation on the rate of capacity deployment (-0,45 * 0,33 = -15 percentage points) still reduces the predicted rate of infrastructure deployment, although this is no longer larger in absolute magnitude than the

positive marginal effect of demand, leading to a predicted net increase in the infrastructure deployment rate of 9 percentage points.

Other independent variables. Our measure of financial constraints, the capital budget of the central government (CAPCOST), is statistically significant and correctly signed, suggesting that countries with looser capital constraints build more capacity. In the sensitivity analysis below, we consider alternate measures of financial constraints. The ratio of imported electricity to total production is statistically significant with a p-value of 0,00, suggesting that the availability of foreign supplies does, on average, dampen the demand for new domestic generation capacity.

Alternative specifications and robustness. Tables 1.2 and 1.3 include results from several additional specifications. Column 2 of each table contains the results from a specification that does not include interaction terms. The effect of industrial representation is once again negative and statistically significant, in accordance with Hypothesis 1. Moreover, VP's lack of statistical significance in this specification is consistent with Hypothesis 3: if the marginal effect of VP is negative where IR is low and positive where IR is high, then the "average" marginal effect of VP—which is what Specification 2 reflects—might well be close to zero.

The results in columns 3 to 5 are based on specifications that respectively use CHECKS3, EXECCON and POLITY as the veto point measure, respectively. The results are qualitatively similar to those in the first column with respect to the Hypotheses 1 and 2, although the effect of VP in Hypothesis 3 is no longer statistically significant, except when IR takes a high value and EXECCON is used as a measure of VP. Additional robustness tests revolve around including various additional and alternative economic influences in our set of independent variables. These include value added in manufacturing, industry and services; urban population percentage; population density; the credit/GDP ratio; and the public sector deficit. In no case is the additional variable statistically significant, nor do the coefficients of central interest change substantially when one of the alternative variables is introduced into the analysis. Additional sensitivity analysis reveals that the reported results are not sensitive to influential data points, as the results are robust to the exclusion of outliers in the dependent variable and the independent variables of theoretical interest.

We also substitute lagged per capita GDP into our core specification as our measure of demand and obtain similar results. Finally, we estimate the core specification for subsamples of OECD and non-OECD countries, and obtain qualitatively similar results to those reported.

Conclusion

A higher level of industrial representation among the consumers of electricity mutes political actors' incentives to satisfy the demands of concentrated geographic interests, labour unions and construction firms to build white elephants, reducing the rate of infrastructure deployment. Veto points that constrain political actors moderate the effect of interest group pressures in the hypothesised manner.

Following recent work in the area of trade, monetary and fiscal policymaking, our results demonstrate the feasibility and importance of combining conceptual perspectives on interest group politics and VP. Where veto points are high are prevalent, interest group pressures have a smaller effect on policy outcomes. The same level of interest group pressure may therefore translate into a different level of "success" in different states or jurisdictions, depending on the formal institutional structure. The analysis further demonstrates the feasibility of capturing international variation in sector-level interest group pressure using readily available economic data.

The effects that we find are statistically, economically and substantively important. Consider the case of Argentina as an example noted by Henisz et.al.(2004) that the unit cost of investment fell from $7,2 million/MW to $1,9 million/MW after the introduction of market-oriented reforms to the electricity sector. This cost reduction translates into annual average savings of $2,74 billion based on the average 517 MW of new capacity that Argentina deployed annually in our dataset.

For purposes of comparison, suppose that Argentina had not undertaken privatisation reforms, but that instead its level of industrial representation had increased from its 1994 value of 41% to 57% (roughly equivalent to the level found in India and Portugal). Given Argentina's current level of VP (POLCON = 0,54), this increase in industrial representation would result in an annual reduction of 722 MW in the rate of generating capacity deployment, saving somewhere up to $5,2 billion annually (depending on whether the unit cost of infrastructure also declined). These calculations would change, of course, for a country with a different institutional environment. For example, a one standard deviation increase in the level of VP (as measured by POLCON) in Argentina's policymaking institutions, to approximately the level of Belgium, would significantly reduce the savings associated with the increased level of industrial representation. In contrast, if the level of VP fell by one standard deviation, to approximately the level of Paraguay, the effect of the increase in industrial representation would rise in magnitude by 1,266 MW annually, saving up $9,1 billion annually.

Despite the strong empirical support for our hypotheses, we also note several limitations that warrant additional cross-national econometric work in this area. First, we are unable to measure the political organisation of the white elephant lobby. Empirical contexts in which national policy debates are easily divisible into consumer versus producer interests, or which pit one region against another, would aid in the further development of the empirical approach that we follow here. Second, our measure of interest group pressure for discipline on SOE infrastructure deployment does not reflect qualitative factors that may affect preferences about redistributive policies, such as dominant national beliefs about the role of the state, especially in energy capacity deployment.

Existing Capacity. The remaining independent variables in the first column of Table 2 are all statistically significant in gauging the effect of the existing capacity level of the infrastructure sector. Finally, our measure of institutional constraints does not take into account the structure of the regulatory apparatus or subnational variation in political and regulatory structures. Despite these limitations, we still derive robust results, consistent with our hypotheses. Better measures should only increase the statistical and economic significance of related findings.

Our ongoing research extends the study of interest group politics and veto players into the more recent period of electricity reform. We examine the timing of privatisation and deregulation of the electricity sector as well as subsequent performance. The insights that we have developed here continue to shape this research agenda. Indeed, we believe that it is only possible to understand the dynamics of the policymaking process through joint consideration of factors discussed in different literatures that cross the disciplinary boundaries of public economics, interest group theories of politics and positive political economy. Failure to do so threatens to introduce omitted variable bias and produce potentially erroneous conclusions. Incorporating multiple policymaking influences better reflects the complexity that investors and political actors face as they interact on a day-to-day basis.