

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.



A CRITICAL EVALUATION OF ROAD PRICING IN SOUTH AFRICA

Duncan Lishman

Supervisor: Associate Professor Anthony Leiman

*Dissertation submitted in partial fulfillment of Masters of Commerce (Applied Economics) degree at
the University of Cape Town*

May 2013

DECLARATION

This dissertation is submitted in partial fulfillment of the degree Masters of Commerce (Applied Economics) at the University of Cape Town.

I know the meaning of plagiarism and declare that all of the work in this dissertation, save for that which is properly acknowledged, is my own. Each contribution to, and quotation in, this dissertation from the work(s) of other people has been attributed, and has been cited and referenced according to the Harvard system of referencing.

I have not allowed, and will not allow anyone to copy my work with the intention of passing it off as their own.

Duncan Lishman

9 May 2013

University of Cape Town

ABSTRACT

Roads are typified by classic symptoms of over-use, with congestion and worn-out surfaces prevalent phenomena. Yet, paradoxically, the infrastructure also suffers from critically inadequate investment. The case is no different for South Africa. In an attempt to provide much-needed finance for road investments, the government's national road agency has introduced numerous toll roads over the last three decades. It is currently in the process of introducing open road tolling on a network of Gauteng's highways. Economic theory provides a rich understanding for pricing road use, particularly with regard to the pricing of externalities. By quantifying these externalities for the South African case, this paper reveals evidence of gross mispricing for road use. Specifically, the magnitude of road freight trucks' external costs indicates that there is an absolute underpricing of road use for these vehicles. On the grounds of the externalities considered in this paper, passenger cars should, in fact, face a toll negligible in comparison to heavy vehicles. That they do not points to massive cross-subsidisation and that the relative price between light and heavy vehicles should be revisited. Appropriate pricing will improve economic efficiency by reducing cross-subsidisation. It will also rationalise the choice of freight modalities in South Africa, with the likely effect that a greater volume of goods will be carried by rail. Despite the welfare gains that the policy offers, one must be cognisant of the distortions that optimal road pricing may have. To assist in achieving second-best optimal pricing, policy proposals are provided for how road use may be practically priced and unrecovered expenses on the road network reduced. It is also proposed that an economic regulator for the road sector be established to develop appropriate pricing principles and provide adequate oversight of the roads sector.

Table of Contents

1. INTRODUCTION	1
2. BACKGROUND TO NATIONAL ROADS IN SOUTH AFRICA	3
2.1. The role of road infrastructure in the economy	3
2.2. History of roads in south africa and their governing institutions.....	5
2.3. Toll roads in south africa	11
3. ECONOMIC THEORY OF ROAD PRICING	19
3.1. Rationale for road pricing.....	19
3.2. Understanding externalities	20
3.3. At what price?.....	21
3.4. Road financing and investment.....	23
3.5. Framework for analysis	24
3.6. On which roads?	27
4. PARTIAL EQUILIBRIUM ANALYSIS OF ROAD PRICING	28
4.1. Applying the theory	28
4.2. The congestive externality	28
4.3. Implications of the congestion toll.....	33
4.4. The road damage externality.....	35
4.5. Quantifying the road damage externality.....	37
4.6. The construction cost externality	45
4.7. Considering sunk costs	48
4.8. Charging for incremental construction costs	48
4.9. Relative mispricing of road use in south africa	49
5. GENERAL EQUILIBRIUM CONSIDERATIONS	51
5.1. From first-best to second-best prices	51
5.2. Diversion to secondary roads.....	52
5.3. Incentive to overload	53
5.4. Pricing in general equilibrium	53
6. A TALE OF TWO CASE STUDIES	57

6.1. Road versus rail – the market for land-based freight transport.....	57
6.2. The gauteng freeway improvement project	62
7. PERTINENT DISCUSSION AND POLICY POINTS	71
7.1. Concerns and considerations	71
7.2. Proposing an economic regulator for roads	75
8. CONCLUSION.....	77
9. REFERENCES	79
10. APPENDICES	89
10.1. Appendix 1 – Major Toll Projects for SANRAL.....	89

List of Figures

Figure 1: National and provincial paved road network length (1938-2001).....	7
Figure 2: Pavement age trend of the national road network.....	13
Figure 3: National road network composition (1972-2011).....	16
Figure 4: Speed-density curve	29
Figure 5: Traffic flow curve.....	29
Figure 6: Travel time - flow curve	30
Figure 7: Deriving the cost curves and optimal toll.....	32
Figure 8: Graphical representation of road damage cost recovery.....	44
Figure 9: Market shares for road and rail freight transport (selected years)	57
Figure 10: Historic road and rail freight volumes, 2002-2012.....	58
Figure 11: Proposed route map for the GFIP.....	65
Figure 12: Proposed N2 Wild Coast Toll Road	90
Figure 13: Proposed N1/N2 Winelands Toll project.....	91
Figure 14: Proposed R300 Ring Toll project.....	91

List of Tables

Table 1: Road expenditure in South Africa (Rm; real 2012 Rands)	9
Table 2: Revenue breakdown for SANRAL (Rm; real 2012 Rands).....	10
Table 3: Revenue breakdown per 100km of SANRAL road (Rm; real 2012 Rands).....	10
Table 4: Selected toll roads currently operational in South Africa	12
Table 5: Potential toll road operations proposed by SANRAL.....	17
Table 6: Forecast SANRAL network composition and length (in kilometres)	18
Table 7: Single-axle equivalency factors	37
Table 8: Comparison table of legal maximum gross vehicle masses (GVM) in 2012.....	38
Table 9: Vehicle specifications of representative road freight trucks	39
Table 10: Road damage by representative vehicles	40
Table 11: Licence fee structure for commercial vehicles	41
Table 12: Truck and trailer population breakdown (December 2012).....	42
Table 13: Weighted average licence fees for representative vehicles.....	42
Table 14: Absolute cost recovery for freight trucks.....	43
Table 15: Proportional cost recovery for freight trucks	44
Table 16: Toll factors between 5 or more axle vehicles and light passenger vehicles.....	49
Table 17: Quasi price basket factor between heavy and light vehicles.....	50
Table 18: Freight cost comparison.....	60
Table 19: Estimated annual road damage costs for representative heavy vehicles.....	61

Table 20: GFIP tolls per kilometre for e-tag holders	66
Table 21: GFIP time-of-day and day-of-week discounts	66
Table 22: E-tag frequent user discounts.....	67
Table 23: Breakdown of GFIP toll collections	69
Table 24: Category breakdown of GFIP toll collections.....	69

ACKNOWLEDGEMENTS

I would like to thank Associate Professor Anthony Leiman for his guidance in the completion of this paper. The generosity with which he gave his time is greatly appreciated, as is the manner in which he shared the depth and breadth of his knowledge

Several other people assisted in a variety of ways: Peter Perkins, for kindly providing the infrastructure data he has accumulated; Nicholas Stiger, for providing handy translating skills of German vehicle regulations; Dalya Levin, for going through the trouble of helping me obtain financial data; and Ryan Lishman, whose non-economics background prompted questions about the theory that I may have otherwise overlooked.

Lastly, I am in debt to the inimitable editing skills of Lara Kerbelker. These are surpassed only by her constant support during the writing of this paper.

Naturally, any errors and omissions are my own.

University of Cape Town

ABBREVIATIONS

GENERAL

AASHTO	American Association of State Highway and Transportation Officials
AVC	Average variable cost
BOT	Build-Operate-Transfer
CBD	Central Business District
CBI	Confederation of British Industry
CCTV	Closed-circuit Television
COSATU	Congress of South African Trade Unions
CPI	Consumer Price Index
CTROM	Comprehensive Toll Road Operations and Maintenance
DBO	Design-Build-Operate
DBSA	Development Bank of Southern African
DoT	Department of Transport
eNaTIS	Electronic National Administration Traffic Information System
ESAL	Equivalent Standard Axle Load
GFB	General Freight Business
GFIP	Gauteng Freeway Improvement Project
GCM	Gross Combination Mass
GVM	Gross Vehicle Mass
IEA	International Energy Agency
JIT	Just-In-Time
LEF	Load Equivalency Factor
MC	Marginal cost

MDS	Market Demand Strategy
MOT	Maintain-Operate-Transfer
NATMAP	National Transport Master Plan
NERSA	National Energy Regulator of South Africa
NPC	National Planning Commission
NTC	National Transport Commission
NRB	National Road Board
ORT	Open Road Tolling
OUTA	Opposition to Urban Tolling Alliance
PCU	Passenger Car Unit
PPF	Production possibilities frontier
PPI	Producer Price Index
PPP	Public-private partnership
PWV	Pretoria-Witwatersrand-Vereeniging
RER	Road Economic Regulator
RFA	Road Freight Association
SAMDM	South African Mechanistic Design Method
SANRAL	South African National Roads Agency Limited
SAR	South African Railways
SAR&H	South African Railway & Harbours
SOC	State-owned Company
SRMC	Short-run marginal cost
TFR	Transnet Freight Rail
TNPA	Transnet National Ports Authority

UNITS OF MEASUREMENT

kg	kilograms
km	kilometres
km/h	kilometres per hour
kN	kilonewtons (where 1kN is approximately 101.972kg-force)
l	litres
Mt	Million tons
Mtpa	Million tons per annum
t	tons (where 1 ton equals 1000kg)

University of Cape Town

1. INTRODUCTION

Toll roads are a topical issue in South Africa. With the impending Gauteng Freeway Improvement Project (GFIP), a sophisticated tolling system for the economic hub of the country, the notion of charging users for access to a previously free resource has raised the ire of consumers, business, lobby groups and politicians alike. In short, the project has generated strong opposition from civil society. Despite this strong opposition from civil society, the Department of Transport and its commercial road agency have pushed ahead with their “user pays” policy, albeit with certain concessions for motorists. In light of the controversy, it is important to understand the broader policy of tolling and its implications for society.

A well-maintained road network of adequate capacity serves an important function in a modern market-based economy. It aids the public, whether commuters in private vehicles or when making use of public transport. It also acts as a system for the movement and delivery of freight goods within the country.

The paper therefore begins by outlining the historical development of the arterial national road system in South Africa, including a view on the current institutional setting. (This section can be omitted without substantial loss of meaning by readers familiar with the local context. The substantive theoretic and analytic sections begin on page 19.) From an institutionally unstable past, South African national road policy has been greatly strengthened under post-apartheid road agency, the South African National Roads Agency Limited (SANRAL). Although toll roads were only introduced in South Africa in the 1980s, they have grown rapidly in length since SANRAL’s inception and are an important source of funding for the agency.

While the market is generally not adept at providing adequate road infrastructure, this does not mean that roads must bear the traditional consequences that are associated with public goods. An important manner in which to ensure efficient use of scarce road infrastructure is to charge users appropriately for use. Economic theory suggests that efficiency is achieved and welfare maximised when price equals the sum of marginal private and external costs, or marginal social cost.

The paper provides an overview of the development of optimal road pricing theory, developing a framework for analysing the negative externalities that arise from road use. Pigouvian taxes close the gap between private and social costs by pricing these external costs, those costs otherwise not borne by the producer.

Non-optimal pricing of road use may result in allocative inefficiencies which distort the value and distribution of social welfare. The negative externalities generated by road freight trucks

are shown to be particularly problematic. For the national allocation of scarce resources to be efficient then prices should reflect the market correctly. From a public policy perspective, hidden taxes or subsidies distort the allocative and productive efficiency of the transport system.

The paper proceeds to evaluate road pricing in general equilibrium. While optimal, first-best tolls follow neatly from partial equilibrium theory, the practical task of deciding whether or not to toll and, if so, on which roads and at what price involves a consideration of the distortions arising from the policy, as well as those that pre-exist in the economy. Good policy ideally achieves an efficiency-enhancing outcome whilst simultaneously addressing equity concerns. By evaluating issues attendant to tolling, one is better positioned to determine socially optimal tolls in general equilibrium, which is where one ultimately finds the role of transport.

The incentives afforded to users and actors in the economy by road pricing are considered, both generally and with specific application. The first of these applications is the general market for freight transport in South Africa, where the past several decades have been marked by a dramatic shift from rail to road for the transport of general freight. This has substantially increased the importance of road haulage to the functioning of South Africa's economy. While a number of reasons have been suggested as proximate causes of this migration, this paper posits that the shift is, at least partly, attributable to the under-pricing of road transport relative to rail. Essentially, road carriers are subsidised by taxpayers and other road users.

The second application returns to the case of the GFIP. The economic importance of the Gauteng region warrants analysis of this project and the evaluation of whether or not these tolls are socially optimal.

Thereafter, the paper concludes by outlining several discussion points with an aim to providing policy proposals that may improve efficiency in the sector.

2. BACKGROUND TO NATIONAL ROADS IN SOUTH AFRICA

2.1. THE ROLE OF ROAD INFRASTRUCTURE IN THE ECONOMY

Transport is an indispensable part of the functioning of the modern economy as “it is a factor in nearly all other economic activities” (Winston, 1985:60). Roads are a critical component of any modern transport system. The movement of passengers and freight between different locations is vital for commercial activity and the functioning of markets. The efficient and effective transport of goods within South Africa, as well as to and from the country’s ports, is vital to facilitate economic growth and development.

Although generally publically-provided, they are an enabler of private enterprise. The centrality of roads has also been recognised in economic theory, where productive infrastructure, including road assets, is recognised as a necessary pre-condition for economic growth. Perkins, Fedderke and Luiz (2005) describe infrastructure as the foundation of the economy, while Jimenez (1995) sees economic infrastructure as existing not for itself but for the purpose of enabling economic activity. Finally, as endogenous growth models have shown, infrastructural expenditures such as roads not only allow productive activities, they also raise the marginal product of existing capital (Barro, 1990). Roads, in particular, can have positive externalities for non-users and the economy in general (Pienaar, 2008:676).

The geographic and topographic layout of South Africa has meant that the country has natural challenges in the movement of freight within its borders, as well as to and from its ports. This problem is caused by three factors. Firstly, the industrial and commercial centre of South Africa’s economy is situated in the landlocked Pretoria-Witwatersrand-Vereeniging (PWV) region, relatively far from any ports. Secondly, this interior region is not accessible by any natural navigable waterways. Thirdly, the east coast of South Africa and its interior are separated by a steep, high escarpment that make the ascent to, and descent from, the hinterland a challenge for the movement of large volumes of freight.

Initially, the problem posed by the South African geography and topography was addressed by an expanding rail network that attempted to link the ports of the country with the industrial centre (Hobart Houghton, 1967). Because of the high escarpment, it was necessary for the railway lines to make numerous turns on their routes. To facilitate these, the authorities introduced the Cape Gauge, a narrower than normal track of 1.067mm (3’6”) (standard gauge measuring 1.435mm [4’8 1/2’’]). This had some of the advantages of narrow gauge, but was

broad enough to allow large loads and reasonable speeds at the time of construction.¹ Investment in the railways was also favoured by the Union government, which formed the state-owned South African Railways (SAR),² the forerunner of the modern day Transnet Freight Rail.

Despite the first-mover advantage of rail, arterial roads that link the commercial and industrial hubs of the country to each other and to the ports have become increasingly important since deregulation of freight transport in South Africa during the 1980s.³ There are certain inherent advantages to road freight transport over rail. Road carriers provide a faster, more flexible service, especially over short distances, and reduce the risks of handling damage and pilferage (Solomon, 1983). These general advantages suggest that road and rail are not substitutes over short hauls but become increasingly close competitors over longer routes. As such, it is necessary that primary consideration be given to the national road system, the network of road infrastructure that spans relatively long geographic distances between the major centres of the country.

In terms of replacement value, the South African road network is undoubtedly one of the country's major assets. With an estimated asset value of R1.7 trillion (National Planning Commission, 2012:167) comes the need to maintain the road network against depreciation. Funding shortages, however, have plagued maintenance schedules, leading to a backlog in road maintenance expenditure estimated at R169 billion in 2012 prices (Paton, 2012). As a result, there has been increased emphasis on user charges as a method of supporting infrastructure investment. Toll roads have therefore become increasingly common on the national road network.

This section provides the institutional and legislative evolution of national roads policy in South Africa. It locates this history in the context of contemporary South Africa and prepares for a detailed analysis of the role and impact of road pricing in the remainder of the paper. Though the institutional administration of the nation's roads has been consolidated under a democratic government, competing demands on the South African fiscus have meant a greater push towards self-financing institutions, amongst which are the national road network.

¹ Indeed, the Cape Gauge persists to this day with the only track laid on the standard gauge being the recently completed high-speed passenger rail service in Gauteng, the Gautrain.

² Part of the South African Railways and Harbours (SAR&H)

³ Even prior to this deregulation, it was noted the importance the road network played in facilitating transportation and thereby economic growth. It was noted that this was, in no small way, because of the lack of navigable rivers in South Africa (Hobart Houghton, 1967).

2.2. HISTORY OF ROADS IN SOUTH AFRICA AND THEIR GOVERNING INSTITUTIONS

2.2.1. The development of national road policy in South Africa (1920s-1997)

In the early days of colonial rule, South Africa's roads were primarily "crude tracks over the veld made by the wagons as they travelled into the interior" (Hobart Houghton, 1967:191). Lord Charles Somerset, governor of the Cape Colony from 1814 to 1826, was the first to oversee the construction of formal roads in the country (De Kock, 1924). Despite the discovery of gold and diamonds in the hinterland and the need to move men, animals and supplies during the Second Anglo-Boer War, the road network of the country remained limited until the advent of motorised transport in the late 19th century (DBSA, 2012:51).^{4,5} Indeed, Van Lingen (1960) reports that in 1905, when the first journey by motor vehicle between Johannesburg and Cape Town took place, it required eleven days over dusty and pot-holed roads and through farm gates.

With the rise in popularity of the motor vehicle, roads took on a heightened significance and a unionised South Africa first promulgated legislation on the matter after the sitting of successive government hearings in the 1920s and 1930s. The first of these was the Holmwood Committee of 1925, which recommended that a national road scheme be established that functioned as a separate but complete arterial system for the country (Department of Transport, 1996).

The Le Roux Commission followed in 1929 and expanded on the work of the Holmwood Committee. Tasked with investigating the impact of nascent motorised road transportation on road and rail services, the Commission's findings were that road transportation in South Africa was "disordered, unrestricted and uncontrolled" (Behrens and Wilkinson, 2001:3).

Two recommendations followed from these findings. The first was for road transport to be brought under the ambit of government regulation, the intention of which may have been to protect the commercial interests of the state-owned railways. This end was achieved through the Motor Carrier Transportation Act 39 of 1930, which introduced a number of Road Transportation Boards to enforce regulations for freight and passenger transport by road

⁴ By way of illustration, it is reported that the majority of the survivors of the 1852 *HMS Birkenhead* shipwreck off Gansbaai opted to settle in the town because it was so difficult to return to Cape Town, despite it being only approximately 160 kilometres away (Havenga, 2007).

⁵ Equipment and supplies for the mining industry were initially transported by means of ox wagons on gravel roads to Kimberley and Johannesburg; later, rail was the first choice after the construction of railway lines from Cape Town to Kimberley and from Cape Town to Johannesburg were completed in 1875 and 1892 respectively (DBSA, 2012:51)

stipulated by the act (Behrens and Wilkinson, 2001).⁶ The act required that operators obtain licences for road haulage from local road transport boards. In practice, however, these licences required the approval of South African Railway and Harbours (SAR&H) (DBSA, 2012:28). This marked the beginning of the rail sector's near seven-decade effective monopoly in freight transport for the country.

The second recommendation was the establishment of a national road board that would be "responsible for the formulation and execution of national road policy, the classification of roads, the location, co-ordination and standardisation of feeder roads and the establishment of a Road Fund" (Department of Transport, 1996:11). Pursuant to this, the National Roads Act 42 of 1935 provided for the establishment of the National Road Board as well as a National Road Fund, a ring-fenced mechanism to fund national road construction and repairs that was financed through a customs duty on imported fuel (Department of Transport, 1996). Construction of the national road network began soon thereafter, although this was interrupted early on by the onset of the Second World War.

Both organisations fell under the Minister of the Interior (Floor, 1985).⁷ While the National Road Board was responsible for the planning and financing of the location, construction and maintenance of a national roads network,⁸ each of the four provinces' Road Departments were responsible for the actual construction and subsequent maintenance of the roads that traversed their borders (Van Lingen, 1960).⁹

Toward the end of 1948, the National Road Board was absorbed by the National Transport Commission (NTC), a body directly under the control of the Department of Transport and established in terms of the Transport (Co-ordination) Act 44 of 1948. The institutional interaction between state and agency, and agency and provinces, appeared to contribute to the Board's demise. With the creation of the Department of Transport in 1943, the autonomy of the Board had come under scrutiny. At the sitting of the Farrar Committee in 1947, the Secretary of Transport lobbied for a re-organisation of the bodies that fell under the department's mandate. Although the Farrar Committee recommended the Board be abolished, the 1945 Page Commission ensured that some distinction was made between national roads policy and the other parts of the road network (Floor, 1985).¹⁰

⁶ These consisted of a Central Road Transportation Board and ten Local Road Transportation Boards, covering ten "proclaimed transport areas" (Behrens and Wilkinson, 2001).

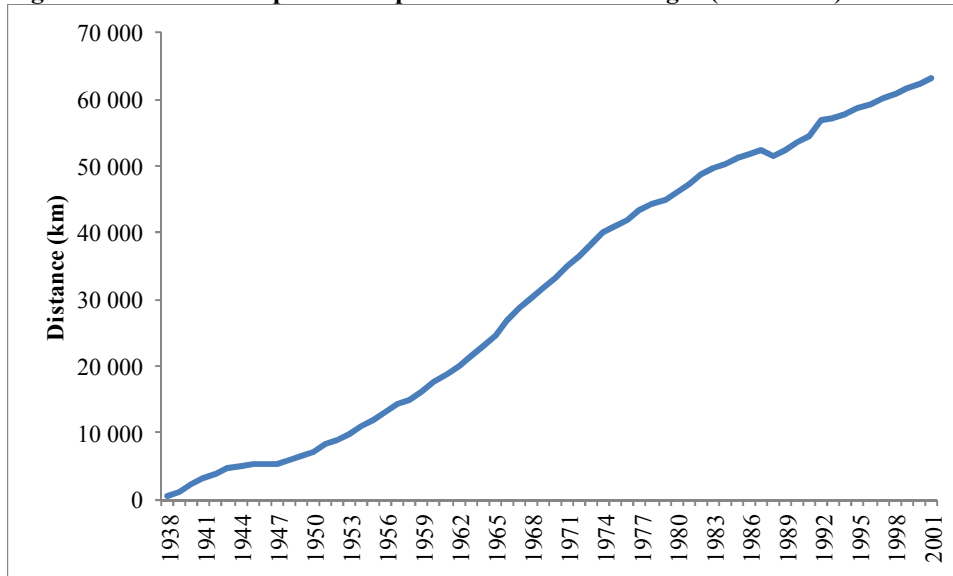
⁷ The Department of Transport only came into existence in 1943 (Floor, 1985).

⁸ In fact, many of the national road routes chosen evolved from historically established routes linking coastal ports with the hinterland (DBSA, 2012:51).

⁹ Until 1994, South Africa was divided into four provincial territories: the Cape Province, Transvaal, Orange Free State and Natal.

¹⁰ The Page Commission was tasked with evaluating the road transport industry and recommending further coordination and regulation of operators. Both its recommendations, and those of the Farrar Committee, led to the Transport (Co-ordination) Act 44 of 1948, which abolished the National Road Board.

Figure 1: National and provincial paved road network length (1938-2001)



Source: Perkins (2003); obtained through a request by the author

Under the new legislation, proposals for national roads were made by an Advisory Committee on Roads. These recommendations, if accepted, were passed on to the NTC for approval. All road construction and maintenance activities of the NTC remained the responsibility of the provinces (Department of Transport, 1996).

By 1961, the NTC had overseen the construction of approximately 90 percent of the national roads originally proposed by the National Road Board. Despite this apparent success, a new National Roads Act 54 of 1971 was adopted to address institutional problems which had hampered road construction in the preceding two decades. This act, and the Transport (Co-ordination) Amendment Act 59 of 1971, increased the independence of the NTC, granting it the sole authority for the planning, construction and maintenance of national roads. The legislative framework also provided for an expanded national roads infrastructure programme, eventually responsible for the modern network of roads that link the major regions of the country (Department of Transport, 1996).

In an effort to fully re-establish the autonomy of the national roads agency, legislation was enacted in 1988 that formed the South African Roads Board (South African Roads Board Act 74 of 1988).¹¹ All national road-related powers of the NTC were consequently transferred to the Roads Board (Department of Transport, 1996).

¹¹ An amendment to the South African Roads Board Act (the South African Roads Board Amendment Act 15 of 1995) further enhanced the body's independence with an expanded board membership that allowed for an even split between members of governmental and non-governmental organisations (Department of Transport, 1996).

2.2.2. National roads policy in contemporary South Africa (1998 – present)

National road infrastructure is currently provided by the South African National Roads Agency Limited (SANRAL), replacing the South African Roads Board. SANRAL was established through the South African National Roads Agency Limited and National Roads Act 7 of 1998. The legislation's enactment has further entrenched the autonomy of the national roads agency and, accordingly, SANRAL is solely responsible for the design, construction, operation, management, and maintenance of the national road network.¹²

SANRAL may be described as a “commercialised public service delivery entity” (Department of Transport, 2006:16). Through its founding act, SANRAL is established to operate on a commercial basis at an arm's length from the government (Department of Transport, 2006). As a state-owned company (SOC), however, SANRAL has the dual goal of being partly self-sustaining, with limited reliance on transfers from the fiscus, while at the same time enhancing social welfare.

2.2.3. Current funding of national roads in South Africa

Given the aggravation over funding for the South Africa road network, the government's intention for the national roads agency to be partly self-financing is of particular interest. The budget for road construction and maintenance has been stretched both prior to as well as after 1994. Funds allocated from the fiscus for roads were, in real terms, less in 1996 than in 1966. This decrease is despite traffic volumes having grown by more than three per cent per annum over the preceding twenty years (Department of Transport, 1996). Sufficient funding for the road network has not been guaranteed since the National Road Fund was dismantled in 1987. New projects have therefore become increasingly difficult to justify in the absence of a dedicated funding source.

Since democracy, the government and the fiscus has faced increased pressure from competing, redistributive social needs.¹³ This pressure may have contributed to the endorsement by Cabinet in 1995 of the “commercialised roads agency” SANRAL with the power to levy user charges (tolls) and, as a result, raise much needed revenue (Department of Transport, 1996:3).

At the same time, SANRAL has had to take control of an increasingly lengthy road system, incorporating formerly provincial roads into the national network. Beginning with around 7

¹² The Constitution of the Republic of South Africa grants responsibility for provincial roads to each of the nine provinces while municipalities are tasked with the provision of basic services including roads.

¹³ See, for instance, Department of Transport (1996); Department of Transport (1999); Department of Transport (2006)

200km in 1998, this network expanded to 16 170 km by the end of 2009 and measured 18 247 km in 2012 (SANRAL, 2009:8; SANRAL, 2009:3).¹⁴

Despite funding pressures, the table below shows that the real cumulative budget for national and provincial roads has grown significantly in recent years. These budgetary increases, however, have mainly been government grants for the construction and upgrade of former provincial roads newly incorporated into the national network. The level of funding is therefore unlikely to continue indefinitely.

Moreover, although funding for road infrastructure and maintenance has increased in recent years, it is difficult to ascertain whether the funding level is adequate for rehabilitation and maintenance needs. Rapid expansion of the network through the incorporation of provincial roads in generally poor conditions means that their elevation to national road standards comes at a significant cost.

Table 1: Road expenditure in South Africa (Rm; real 2012 Rands)¹⁵

	2003	2004	2005	2006	2007	2008	2009	2010
Provincial	R 7 178	R 8 395	R 9 193	R 10 531	R 11 351	R 12 073	R 13 808	R 14 366
SANRAL	R 1 643	R 1 871	R 2 078	R 2 425	R 4 218	R 6 959	R 14 125	R 19 813
Total	R 8 820	R 10 266	R 11 272	R 12 956	R 15 569	R 19 032	R 27 933	R 34 179

Source: DBSA (2012:55)

The result of the high cost of road infrastructure in the face of constrained funding is evidenced in two ways in South Africa. Firstly, there has been significant underinvestment in the road network. While this has been apparent since the 1960s,¹⁶ the fiscus has been increasingly ill-equipped to adequately fund road construction and maintenance. (The current maintenance backlog is evidence of this.) Secondly, there has been an increased use of toll roads to finance construction and maintenance of new and existing sections of the road network.

As a result, the current national road network consists of two types of roads: tolled (19 percent of the network) and open or untolled (81 percent of the network). SANRAL's founding legislation requires a clear separation in the funding of these roads with no cross-subsidisation permitted between the two. While toll roads are funded by user charges, non-toll roads are funded by allocations from the fiscus. The operation of toll roads, and the funds collected from their use, is the responsibility of either SANRAL or an appointed concessionaire (SANRAL, 2009:15).

¹⁴ At 31 March 2012 (SANRAL's financial year end) 16 170km were incorporated as national roads. Subsequent to this date, an additional 2 077km of formerly provincial roads have been incorporated (SANRAL, 2012c:98).

¹⁵ All Rand figures have been adjusted for changes in the civil engineering plant index of the Producer Price Index (PPI), rather than the Consumer Price Index (CPI). Changes in the former more accurately estimate the real value of funding.

¹⁶ Admittedly, the 1960s were characterised by major infrastructure spending on highway development. Regardless, the recurring funding for maintenance purposes has not kept track with the growing size of the network.

Since the early 2000s, when toll income made up nearly half of SANRAL's funding, the share of toll income has fallen as the government has increased transfers to the agency.¹⁷ Absolute figures suggest that government grants have provided better funding for roads in recent years.

Table 2: Breakdown of SANRAL's income by source (Rm; real 2012 Rands)

Rm	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Government grant - revenue portion	R 880	R 930	R 833	R 1 770	R 1 919	R 1 847	R 2 095	R 1 967	R 2 403	R 3 455	R 4 640
Government grant - capital portion	R 319	R 625	R 953	R 309	R 609	R 1 163	R 1 755	R 2 398	R 3 394	R 3 792	R 9 591
Total government grant	R 1 200	R 1 554	R 1 786	R 2 078	R 2 528	R 3 010	R 3 850	R 4 365	R 5 797	R 7 247	R 14 231
Toll income	R 649	R 737	R 996	R 1 185	R 1 265	R 1 342	R 1 313	R 1 262	R 1 481	R 1 764	R 1 900
TOTAL INCOME	R 1 529	R 1 667	R 1 828	R 2 955	R 3 184	R 3 189	R 3 408	R 3 229	R 3 884	R 5 219	R 6 539
<i>Grant revenue as proportion of income</i>	58%	56%	46%	60%	60%	58%	61%	61%	62%	66%	71%
<i>Toll revenue as proportion of income</i>	42%	44%	54%	40%	40%	42%	39%	39%	38%	34%	29%

Source: SANRAL annual reports (2000-2012)

Note: SANRAL's revenue figures do not include toll revenues raised through its three PPPs.

Given that the length of the network has grown tremendously, it is instructive to analyse that revenue adjusted for the number of kilometres of road under SANRAL's management. From Table 3, it can be seen that toll roads are almost three times better funded than non-toll roads on a distance-adjusted basis. This despite the fact that funding for the same-length roads without tolls has increased in real terms by 76 percent in the past 11 years, while toll roads have actually experienced a slight decline in funding.

Including the provision of capital grants from government indicates a closer funding scenario between the two road types, though only really similar in the 2012 figure. However, this figure includes a once-off R5.75 billion disbursement to reduce the outstanding capital debt incurred by SANRAL on the GFIP and is therefore unlikely to be repeated. This suggests that, even when incorporating capital expenditure amounts, non-toll roads have approximately half the available funding as compared to toll roads.

Table 3: Breakdown of SANRAL's income per 100km of road (Rm; real 2012 Rands)

Rm	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Government grant (total) per 100km	R 23	R 29	R 24	R 23	R 19	R 22	R 28	R 33	R 44	R 55	R 107
Government grant (revenue) per 100km	R 17	R 17	R 11	R 20	R 14	R 13	R 15	R 15	R 18	R 26	R 35
Toll revenue per 100km	R 109	R 124	R 167	R 199	R 114	R 121	R 118	R 69	R 81	R 96	R 104
Average recurring revenue per 100km	R 21	R 23	R 20	R 27	R 20	R 20	R 21	R 20	R 24	R 32	R 40

Source: Author's own calculations

Note 1: Total government grants include money provided from the fiscus for capital expenditure on new road infrastructure

Note 2: These figures are only approximate as, given data constraints, they do not match exact changes in network length and the declaration of sections as toll roads. They are therefore only a guideline for the average funding size.

¹⁷ Note that this toll income is not a completely accurate reflection of collections from the entire system as BOT toll revenues are not passed through SANRAL's income statement.

It is clear that an increasing network length necessitates greater emphasis on the mode of funding. Tolling has been an important factor in delivering a good quality road network and, as can be seen, provides far better funding for its stretches of road than government funding.

2.3. TOLL ROADS IN SOUTH AFRICA

2.3.1. Contextual background

While toll roads have become increasingly common in South Africa in recent years, they are not new to the country. As early as the 1840s, charges for the use of roads and bridges are recorded. A Central Board of Commissioners for Public Roads was created by the colonial secretary of the Cape Colony, John Montagu, in 1844 to oversee the construction of highways, bridges and mountain passes in the region. Furthermore, the Board was empowered to levy a toll on users of the roads and bridges (De Kock, 1924). An example of such a toll facility was that of Garcia Pass, a 20km stretch of road between Riversdale and Ladismith which charged three pennies for a wagon and one penny for a bicycle (Maluleka, 2008). After the unification of South Africa, however, all road financing was provided through the National Road Fund and government grants, and the concept of road user charges was seemingly shelved.

Toll roads in contemporary South Africa were first considered in the 1960s and 1970s as a method to finance the construction of the country's urban freeway systems. In 1977, the authorities decided against such a scheme, citing obstacles to its viability, including high administrative costs, the required high traffic volumes and consumer resistance. However, a governmental mission to the Far East in 1980 led to a reversal of this decision. The government deliberated over the inflationary effects of a fuel levy increase versus selected tolling schemes, eventually opting for tolls for their more localised effect on prices.¹⁸ As a result, Parliament adopted the Second National Roads Amendment Act 79 of 1983, providing the NTC with the authority to implement toll roads amid the financial difficulties faced by the National Road Fund (Floor, 1985). The new legislation provided for tolls to fund new roads, or road improvements on stretches where an alternative route existed, i.e. initially traffic deviation was presumed and built into the system.

The first modern toll road followed in June 1984 when the Tsitsikamma Toll Road was opened to traffic. Conceptualised by the National Road Board in 1938, the route was ten kilometres shorter than its predecessor and avoided previously "treacherous" passes through the use of concrete-arched and concrete strut-framed bridges (Floor, 1985:79).

¹⁸ At the time, the country was battling with a persistently high rate of inflation largely driven by rising oil prices. This gives some context for the debate.

Subsequent to the introduction of toll roads, the NTC found the main advantage of tolling a section of road was the ability to borrow funds for construction against the expected income from user charges. This income was expected to sufficiently cover both the financing component (interest and principal repayment) and the maintenance and operation costs (Floor, 1985). With the introduction of the South African Roads Board Act No 74 of 1988, the duties for national roads and implementing tolls on this road network vested with the South African Roads Board.

2.3.2. Toll roads in democratic South Africa

The subsequent transfer of balance sheet, powers and responsibilities to SANRAL meant the national roads agency also inherited the power to levy user charges on those parts of the national road network it deemed toll roads. The legislative framework provides SANRAL with the sole authority to charge for use of national roads, repealing all previous legislation.¹⁹

Toll roads have become an increasingly common feature of the South African road network since the Tsitsikamma Toll Road was opened, covering an increasing length of the national network and an associated higher volume of traffic (Leiman, 2003).

Table 4: Selected toll roads currently operational in South Africa

Name	Road	Distance (km)	Tolling start date	Operator	Contract form
N3 Toll Road	N3	418	November 1999	N3TC	PPP BOT (30 year)
N2 Tsitsikamma	N2	53	June 1984	SANRAL	CTROM-operated and -maintained
N4 West (Platinum Highway)	N4	380	August 2001	Bakwena	PPP BOT (30 year)
N4 East (Maputo Development Corridor)	N4	504	February 1998	TRAC	PPP BOT (30 year)
Chapman's Peak	Chapman's Peak Drive	9	2003	Entilini/Tolcon	PPP BOT (30 year)
GFIP Phase 1	Various in Gauteng	185	Pending	SANRAL	CTROM-operated and -maintained

Source: Various SANRAL annual reports; Chapman's Peak Drive website

Note: The N2 Tsitsikamma toll road was extended between 2007 and 2010 from 28km to 53km, with toll fees increased accordingly.

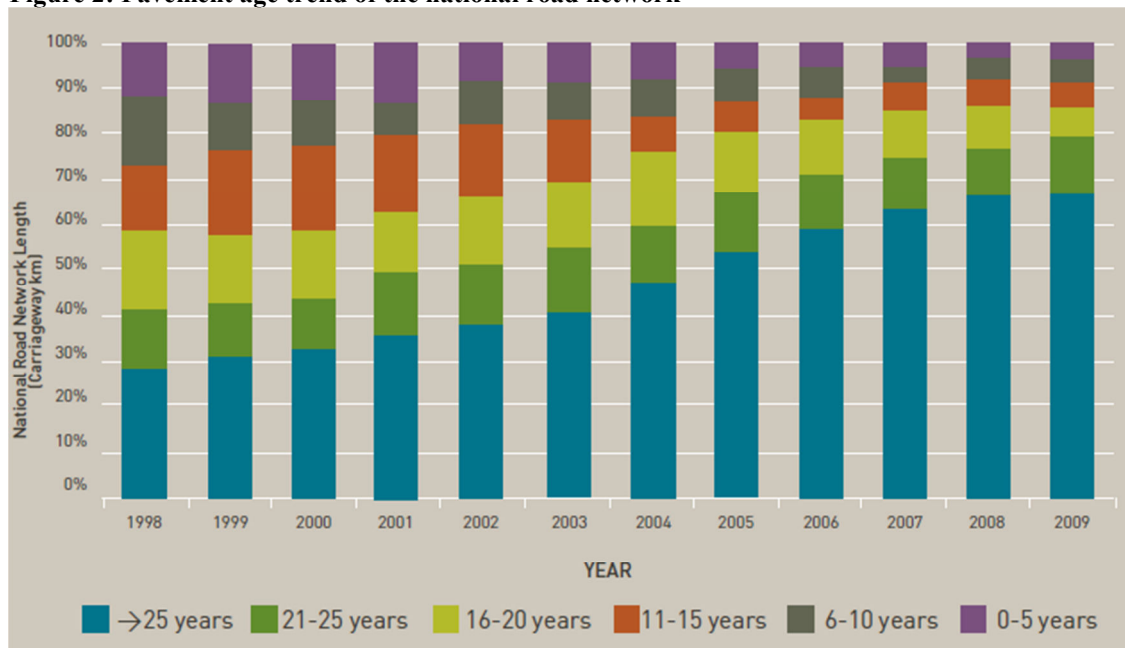
This paper posits that there are five proximate factors that have given rise to the proliferation of toll roads.

The first, and primary, reason is financing the construction and maintenance of national roads. The Department of Transport's "long term policy for roads funding will focus on increasing investment in roads... Increased investment will be achieved through, [*inter alia*, a] review and improvement of tolling as a strategic approach for financing appropriate sections of the national road network" (Department of Transport, 2005:16).

¹⁹ Indeed, the only non-SANRAL toll road in the country is that of Chapman's Peak Drive in the Western Cape. The enacting legislation, the Western Cape Toll Roads Act 11 of 1999, more generally empowers the provincial minister responsible for transport to proclaim toll roads and to enter into agreements for the construction and/or operation of those toll roads (Department of Transport, 2006).

Toll roads provide much needed finance to a road network that is largely older than its design life. SANRAL stated that, in 2009, more than half of South Africa’s non-toll road network exceeded its original lifespan. As the Development Bank of South Africa (DBSA) points out, South Africa’s road network has “deteriorated due to over-utilisation and under-investment” (DBSA, 2012:47). The continued expansion of the network under SANRAL’s management through the incorporation of provincial roads, generally in a poor condition, has aggravated such funding pressures.²⁰

Figure 2: Pavement age trend of the national road network



Source: SANRAL (2009:41)

Where funding provided by the National Treasury is insufficient to adequately maintain the entire national road network, the declaration of sections of the network as self-funding toll roads provide a dual benefit. The tolls allow for the upgrade and maintenance of that section of road to acceptable standards. Funds from National Treasury can then be spread across the smaller remaining national road network. The incentives of government and its agency are therefore aligned to introduce toll roads on the national network wherever viable.

Secondly, funding pressures also suggest there is an intuitive appeal for the government of a user-pays system. Infrastructure is provided for the benefit of an identifiable group of users and prices are set to recoup the expenses of its provision.

Thirdly, SANRAL possesses a less encumbered regulatory ability to implement tolls. The Second National Roads Amendment Act 79 of 1983 precluded the NTC from declaring any

²⁰ Although the figure does not provide absolute figures, it is likely that the drive force behind the pattern in Figure 2 is the incorporation of former provincial roads.

portion of a national road as a toll road unless an alternative road was available to users via which both destinations, and destinations in between, could be reached.²¹ This section has been done away with in the new legislation, granting the roads agency with a wider scope to declare toll roads. SANRAL's founding legislation also allows for the creation of toll roads through "unsolicited bids", whereby private consortia may provide design, build and operate (DBO) plans for new sections of national road or build, operate and transfer (BOT) bids to SANRAL for existing sections of the network. In both instances, bids may be submitted to the agency without it having advertised for such bids.

Fourthly, SANRAL may face a skills constraint in terms of its engineering and project management capabilities. Though it recognises the national road network as essential infrastructure, it is relinquishing responsibility by outsourcing toll road maintenance and operations to private contractors to alleviate this constraint.

Finally, the Department of Transport sees user charges as bringing economic efficiency to the transport system. The democratic government's initial transport policy paper, "Moving South Africa", suggests that a framework be established that adopts "the principle that those who cause the externality costs should pay for them ... [and that these costs] should be 'internalised' back into the transport system" (Department of Transport, 1999:9). The internalisation of these road user costs would help restore "economic rationality" to the system (see, for instance, Department of Transport, 1999; Department of Transport, 2006). The policy paper goes on to espouse sound economic principles:

Road-use should be priced to fully recover the costs of infrastructure provision and maintenance, as well as externalities. This is necessary to restore economic logic to the system. Such a step will ensure long-term sustainability, reduce the negative externality effects, and create a self-supporting system that sends the right price signals for using roads. Correct price signals for different environments (e.g. urban and rural) will help to prevent congestion and pollution. (Department of Transport, 1999:68)

Presumably this policy is embodied when users of the transport system face the correct price signals for their use. Toll roads, therefore, provide the opportunity for the cost of the externalities generated by road use to be charged in full to the appropriate user. It will be shown, however, that despite policy pronouncements this has not occurred in the past fourteen years.

²¹ Section 9(3) of the Act.

2.3.3. Implementation of tolls

SANRAL may implement toll roads on the national network in two ways. It may either finance and operate its own toll roads or, alternatively, allow private operators to do so through concessions.

In the first instance, SANRAL can either borrow against its balance sheet to finance the construction of the toll road or it can enter into an agreement in terms of which a private contractor finances the project. In principle, neither method requires a direct injection from the fiscus.

Though SANRAL may collect tolls itself, it developed the Comprehensive Toll Road Operations and Maintenance (CTROM) contract model during the early 2000s for the outsourcing of operations and maintenance of its toll facilities. These are tendered out as single contracts in which the operating company takes responsibility for all aspects of operation and maintenance of the toll route (SANRAL, 2001:15). By 2002, this model had been implemented on all of the agency's toll routes (SANRAL, 2002:17).

Where SANRAL uses the private sector for toll roads through a build, operate, transfer (BOT) scheme, this may either be through a competitive tender process or via an unsolicited bid. With the former, SANRAL selects a section of road to toll and then tenders this out to prospective bidders. Since its inception, SANRAL has successfully developed three public-private partnerships (PPPs) for the development and operation of toll roads in this manner. These are along the N3 (Gauteng-Durban), N4 Platinum Highway (west from Gauteng through Rustenburg to Zeerust and the Botswana border), and the N4 Maputo Development Corridor (Gauteng to Maputo).

While competitive tendering procedures are still used by SANRAL, the agency also allows for unsolicited bids. The concept of unsolicited bids is outlined by Leiman:

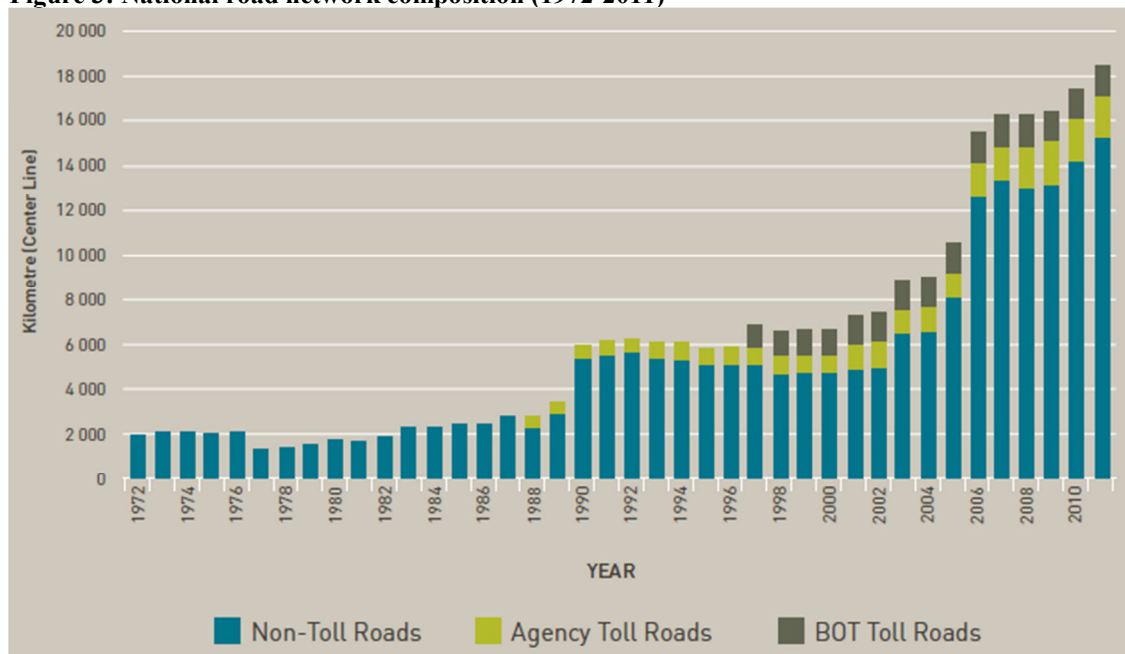
Typically a consortium of contractors, operators and banks approaches SANRA[L] with an offer to construct a new road or to take over the management and maintenance of an existing one, on a [build, operate, transfer] contract. SANRA[L] considers the proposal and may accept or reject it. If accepted, other consortia may present counter-bids within 60 days. (Leiman, 2003:126)

This policy was introduced in 1998 and, by 2003, seven unsolicited bids had been received by SANRAL had been given "Scheme Developer Status". Though no unsolicited bids have been

implemented a decade later,²² three of SANRAL’s major planned construction projects are unsolicited bids. These are the R300 Ring Toll Road, the N1/N2 Winelands and N2 Wild Coast toll roads, with further information provided in the appendix.

This, however, has not inhibited the growth in length of toll roads over the last fifteen years, as shown in the figure below. SANRAL has recently completed a flurry of toll road projects, including phase 1 of the Gauteng Freeway Improvement Project (GFIP), the N17 from Springs to Ermelo, the R30 from Bloemfontein to Ermelo, the N2 Tsitsikamma Toll Road Extension, and the N2 Dube Trade Port (King Shaka International Airport).

Figure 3: National road network composition (1972-2011)



Source: SANRAL (2009:40)

Practically, SANRAL and toll consortiums collect tolls through toll plazas, though open road tolling (ORT) is used on several stretches of toll roads, including the new GFIP (exclusively ORT) and certain lanes of the N3 and N4. No tolls are charged on a new toll road until “initial construction” has been completed (SANRAL, 2009:32). For operational toll roads, prices are generally submitted by the toll operator to SANRAL for approval, and are then vetted by the Minister of Transport.

2.3.4. Planned toll road projects going forward

SANRAL has developed a list of projects that it considers eligible for tolling going forward, provided in the table below. Together with the 3 120km currently tolled, this will take the size of the toll road network to nearly three times its current length.

²² SANRAL (2009:36) indicates all PPPs thus far have been SANRAL-initiated.

Table 5: Potential toll road operations proposed by SANRAL

Road	Section	Approximate distance
		(km)
N12	Johannesburg to Three Sisters	901
N1	Kroonstad to Three Sisters	671
N2	Wild Coast Toll Road	540*
R33	N4 Nelspruit to Polokwane	524
N2	Richards Bay to Ermelo	436
N11	Ladysmith to Middelburg	349
R72	East London to Alexandria	232
N5	Winburg to Harrismith	224
-	GFIP Phase 2	223
N1/N2	Winelands Toll Road	175*
R33	Piet Retief to N4 Belfast	172
R33/R510	Nylstroom to Lephale	166
N8	Kimberley to Bloemfontein	150
N8	Bloemfontein to Maseru	136
N4	Belfast to N11	104
N3	Marianhill extension	86
R300	Cape Town Ring Road	68*
N12	Edenvale to Daveyton	27
R512	PWV3/Magalies Extension	9
N1	Huguenot second tunnel	5
Planned length for tolling		5198
Current length tolled		3120
Total length of network to be tolled		8318

Source: SANRAL (2009:16)

Note: Routes marked with a * refer to unsolicited bid proposals.

This list was set out in SANRAL's 2009 "Declaration of Intent" for the 2009/2010-2012/2013 period. In the interim, the political environment for toll roads has changed somewhat, with the strong opposition to the GFIP from civil society and the general public, creating uncertainty over when a concerted programme of toll road construction will resume. However, the stated policy of the government remains the user pays principle. This is reaffirmed in the National Development Plan, adopted by Cabinet, and SANRAL stated in January 2013 that "government policy still supports the user-pays principle and SANRAL will selectively use tolling to fund [road] upgrades" (Munshi, 2012:41).

2.3.5. Forecast size of the national road network

Although the eventual national road network size was originally projected to be 19 077km in length, recent indications are that the revised network length will measure between 32 000km

and 38 000km (SANRAL, 2012b:12; SANRAL, 2012c:16; DBSA, 2012:47). Compared to this, the remaining provincial road network will have a surfaced length of 31 474km.

From Table 6, this figure is composed of the current SANRAL network plus the “remaining strategic network”, those portions of the declared national roads not yet under SANRAL’s management, and the “primary network”, which predominantly forms part of the provincial arterial road network.

Table 6: Forecast SANRAL network composition and length (in kilometres)

Province	SANRAL (current)	Remaining strategic network	Primary network	Total
Eastern Cape	2433	227	2366	5026
Free State	1592	479	2460	4531
Gauteng	617	249	840	1706
KwaZulu-Natal	1422	804	1405	3631
Limpopo	1922	158	963	3043
Mpumalanga	2300	59	1552	3911
Northern Cape	3114	12	1219	4345
North-West	1303	320	1253	2876
Western Cape	1467	599	1748	3814
Total	16170	2907	13806	32883

Source: SANRAL (2012b)

It is unclear what portion of this expanded network SANRAL sees as suitable for tolling. However, indications from Table 5 suggest that a potential 25 percent of the total network could be tolled. Crucially, these roads are tolled because of the traffic volumes they carry, which generally indicates their corresponding economic and social importance. The manner in which use is priced therefore has wider efficiency and welfare implications, and a clear understanding of pricing principles is therefore required.

3. ECONOMIC THEORY OF ROAD PRICING

3.1. RATIONALE FOR ROAD PRICING

Roads may be considered quasi-public goods (Mohring and Boyd, 1971). They are so defined because though they behave as public goods under certain circumstances, there is a degree of excludability (people can be excluded from use of the resource) and they may also be rivalrous once some scarcity is applied (congestion may arise).²³ As quasi-public goods, roads are generally provided by the government rather than the market, as no competitive market exists to efficiently allocate the necessary resources (Winston, 1985).²⁴ In the absence of a market outcome, it is important to establish how the use of road infrastructure should be priced so as to send the correct signals to ensure efficient use in the present and sufficient investment in the future. Unpriced roads tend to exhibit characteristics typical of command economy goods and services: in the short run, congestion as motor vehicles queue to make use of limited road space; and in the long run, a chronic lack of funding for investment (Hau, 1998).²⁵

There is a remarkable consensus among economists that society would benefit from optimally priced road use. There is, however, less of a consensus about exactly what such a price should entail. There are four classic policy interventions for road pricing: fees (for example, vehicle licensing), which are generally lump-sum taxes and therefore have little behavioural effect at the margin; fuel taxes, where total payment varies directly with road use and therefore may affect behaviour; tolls; and congestion charges (time-based tolls to reduce peak-hour congestion). For the purposes of this paper, these charges will be referred to collectively as the quasi-market price.

Standard economic theory posits that users must face a price for the efficient use of any scarce resource.²⁶ On the demand side, those who attain benefits from the use of the resource at a value equal to or greater than the price will enjoy continued use. Others will adjust consumption downwards, either substituting for an alternative or foregoing the good entirely. For suppliers, prices act as important signals for production adjustment in the short run and investment changes in the long run.

²³ As it will be shown, this rivalrous nature may be mitigated by appropriate user pricing.

²⁴ Left to market forces, market failure ensues. Roads are generally publically provided goods; it is highly unlikely that societies would have the road networks they do if it were not for government spending, even those roads for which private provision would be “profitable”

²⁵ It is often suggested that the short run phenomenon of heavy congestion may be alleviated by expanding and upgrading the road network. Though this is true in the short-run, the operation of Downs’ Law on highways in the long run means that latent demand emerges to meet maximum road capacity and the system returns to its previous (congested peak-hour) state (Downs, 1962).

Though this phenomenon is frequently cited in the literature, the opposite could easily be argued. The development theory of unbalanced growth, shaped by economists like Albert Hirschman, suggests that investment in infrastructure such as roads spur economic development. While one may observe empty highways initially, their construction and existence induce the economy to expand.

²⁶ More specifically, they must face the “correct” price

Where a resource's price is not optimal, common outcomes are inefficient use in the short run and inappropriate investment in the long run. Without correct pricing, roads share these short and long run problems.

Pricing road use would improve the state of road transportation and enhance economic efficiency by dealing with these issues. The government may do this by setting quasi-market prices that require users to bear the costs they impose on others (Pienaar and Nel, 2009:46). This road price would increase social welfare because it would confront every road user with the full social costs of each kilometre travelled (Verhoef, 2008:10). In the absence of road pricing to correct for externalities, one has the "socially excessive consumption of road trips" (Verhoef, 2008:7). It is the effective and optimal pricing of these externalities that is necessary to improve welfare and achieve economic efficiency.

At the core of the notion of road pricing is a three-part decision process by the policy-maker. The first decision is whether to use road pricing or not. This addresses the rationale for the use of economic theory. Secondly, if the answer to the first question is affirmative, one must then formulate a price at which to toll. Following from the theory, an optimal price exists such that economic efficiency is maximised, at least in a partial equilibrium setting. Thirdly, it must be determined on which roads prices should be charged.

3.2. UNDERSTANDING EXTERNALITIES

Most externalities of relevance for public policy relate to public goods (Baumol and Oates (1988:14). For our purposes, these externalities are in relation to quasi-public goods.²⁷

A distinction may be drawn between externalities that are pecuniary and those that are technological (Baumol and Oates, 1988:29). Pecuniary externalities are those that manifest themselves through changes in the prices of input or outputs. Such changes influence the values of real resources and therefore have the effect of moving the production set along the production possibility frontier (PPF). Although the resource allocation is altered, there is indifference between the two states.

In contrast, technological externalities shift the functions relating resource quantities to output or utility levels (Baumol and Oates, 1988:30). While they may also impact on prices and real resource values, the important distinction is that a negative technical externality moves the PPF inwards rather than merely changing the economy's position on it. While pecuniary externalities have income distributional but not efficiency effects, technological externalities

²⁷ To be sure, Mohring and Boyd (1971) indicate that the set of "pure" public goods is empty.

reduce efficiency. The externalities that are of relevance in the context of road use are technological in nature.

The standard Pigouvian result for the pricing of a negative (technological) externality, in a competitive economy, calls for a tax per unit on the externality-generating activity equal to its marginal external cost (Baumol and Oates, 1988:55).²⁸ While Pigouvian taxes are optimal only under perfect competition, roads are a quasi-public good not subject to competitive market outcomes. Instead, they are generally provided by the government, a monopolist in supply. Buchanan (1969) and Barnett (1980) commented that Pigouvian taxes under a monopolist may not lead to optimality and, in fact, could reduce welfare. If a profit maximising monopolist has already restricted his output below competitive levels, the additional output reduction resulting from the externality charge may result in the aggregate social welfare gain from externality-pricing being outweighed by the welfare loss from lower levels of production (road use) (Baumol and Oates, 1988:79-80).

The existence of such distortions in the economy suggests that truly optimal tolls will be impossible to find. Policy makers should aim for tolls that reflect 'second-best' conditions. This will be explored in section 1.

An additional caveat is that externalities in road use may imply certain non-convexities (i.e. multiple apparent equilibria). This phenomenon may lead to a situation where, despite optimal pricing, one does not actually maximise welfare. There is evidence that there may be economies of scale present in road widening and thickening (Winston, 1991) and non-convexities may therefore be problematic. However, it will be argued that the existence of diseconomies of scope in road construction may act to cancel out this effect.

3.3. AT WHAT PRICE?

While Arthur Pigou (1920) was the first person to introduce road use in economic terms,²⁹ the modern analytical framework for road pricing owes much of its development to the work of Mohring and Harwitz (1962).³⁰ William Vickrey contributed to the discourse, identifying early on that it was not the case that fuel and vehicle taxes sufficiently covered the cost of road use (Vickrey, 1963; Vickrey, 1969). More recently, a thorough mathematical exposition on the components of road pricing has been conducted by Newbery (1989) while Hau (1992; 1998)

²⁸ The source of an externality is typically found in the absence of adequately defined property rights (Baumol and Oates, 1988:26). If such rights were allocated in full, and given costless bargaining the optimal outcome would emerge regardless of the original allocation of property rights (again, only in perfect competition).

²⁹ Pigou's initial work on roads in *Welfare Economics* was heavily criticised by contemporary Frank Knight, to such an extent that Pigou withdrew the analysis from future editions (Hau, 1998:42).

³⁰ Although Pigou discussed the economics of road use in the 1920s, the literature only really begins in the 1960s with Mohring and Harwitz (1962).

has provided a rigorous graphical derivation. Verhoef and Mohring (2009) recently produced a paper on the implications of theoretical road pricing for practical policy decisions.

At its core, the implementation of optimal road pricing takes the view that road users generate negative externalities for which they do not always face a price. Economic efficiency can be achieved through the use of short-run marginal cost (SRMC) pricing to induce users to internalise the cost of the externalities they generate.³¹

The SRMC price (or Pigouvian tax) should equal the difference between the marginal social costs and marginal private costs of road use.³² Broadly, the costs incurred in road use can be separated into private and external costs. Private costs are directly incurred by the user in the form of fuel, maintenance, depreciation of the vehicle, and other ancillary costs such as the monetary value of time. External costs are those costs that are not incurred directly by the user but, instead, are borne by third parties or society. Internalising the externality implies that the negative effects on society are taken into account by appropriately pricing the external costs generated by the additional user as he/she joins the road. When these external costs are internalised, the user therefore faces the correct price of the decision to make use of the road.³³

This differs from a policy instrument like a lump-sum tax on road use that may act to reduce demand, not because the behaviour of users has changed but rather because some fraction of users has exited the market. Given the positive externalities that roads offer the economy such a pricing mechanism is problematic. Road use is the basis of much economic activity and social welfare. Too high a price means too little use of roads and a decline in national well-being. Too low a price may diminish social welfare and lead to allocative inefficiency (observable in the road system when congestion is present).

An alternative to SRMC pricing is long-run marginal cost pricing, which may be advocated for the purposes of capital investment recovery. However, Hau (1998:53) argues that SRMC is the pricing rule to use regardless of timeline. Only an optimally built road would have the long-run marginal cost price send the correct price signal. An underbuilt road (one where inadequate capacity exists, such that congestion occurs) would set too low a price, while an overbuilt road (one where excess capacity exists for a given level of traffic) would overprice use. SRMC pricing is preferable as the price can be varied according to demand, and therefore capacity, patterns.

³¹ The theory of marginal-cost pricing has been critiqued by Graaff (1968:154-155), where he argues that the conditions under which price equals marginal cost "are so restrictive that they are unlikely to be satisfied in practice". As a useful starting point for analysis, he criticises marginal cost as being no more a useful starting point than the value zero.

³² It will be shown that under the congestion pricing model presented, marginal private cost and average cost are analogous.

³³ Note, however, that only under perfect competition is the optimal tax equal to the marginal external cost.

3.4. ROAD FINANCING AND INVESTMENT

Importantly, the design of road tolls must not be with the sole purpose of raising revenue (be it for investment, cost recovery or otherwise). The market mechanism relies on prices to allocate scarce resources. If it is to serve this allocative function effectively, a road toll should not be set with the sole purpose of financing infrastructure, as practiced by SANRAL. Vickery argues that:

[t]he delusion still persists that the primary role of pricing should always be that of financing the service rather than that of promoting economy in its use. In practice there are many alternative ways of financing; but no device can function quite as effectively and smoothly as a properly designed price structure in controlling use and providing a guide to the efficient deployment of capital. (Vickery, 1963: 455)

Ideally, pricing should aim to have some tangible effects on the magnitude and composition of traffic. If it does not, the economy will end up “with a socioeconomic deficit equal to the system investment and operations costs” (Johansson-Stenman and Sterner, 1998:155).³⁴

By taxing externalities rather than productive economic activities, road pricing is a good principle for public financing (Atkinson and Stiglitz, 1980). It provides a source of funding for a public good with a record of underinvestment, across countries and time, without the need to explicitly raise revenue to fund the underlying infrastructure.

The literature suggests that an optimally designed road, in terms of capacity and toll, can be self-financing if constant economies of scale are present and road capacity can be changed in continuous increments (Mohring and Harwitz, 1962; Mohring, 1970). Provided that SRMC pricing is applied and that quasi-rents (or losses) are used to adjust the capital stock, then the total revenue garnered by marginal cost pricing will be sufficient to cover total costs. (This is a novel finding given that the sum of marginal cost equals total variable cost and not total cost.) However, if there are decreasing returns to scale the road will earn a surplus, while if there are increasing returns to scale it will operate at a deficit (Mohring and Harwitz, 1962; cited in Winston, 1985).

Economies of scope cause the total cost of producing multiple products jointly to be less than the cost of producing each product separately (Winston, 1985:64). There is evidence for the existence of diseconomies of scope in road construction. Indeed, it may be cheaper to construct separate road systems for trucks and passenger vehicles. This is because as the road is widened to accommodate more traffic, the cost of the additional thickness required to handle

³⁴ The obvious assumption here is a road tolling system with non-negligible investment and operations costs.

heavy vehicles increases as all lanes on a road will necessarily be built to the same thickness. The cost of “producing” passenger traffic is therefore higher when it has to accommodate heavy freight traffic (Winston, 1991).

In reality, it may be the case that the economies of scale available in road transport (widening lanes, increasing thickness) are more or less offset by these diseconomies of scope, such that the two net off for constant returns (Small, Winston and Evans, 1989; Winston, 1991). Even if it is the case that economies of scale dominate, charging users for the costs they impose on society would go some way to providing capital and maintenance funding for the road network.

Regarding the case where revenues exceed infrastructure costs, Keeler and Small state that:

objections have been raised to short-run peak [congestion] tolls because they will result in excessive profits for road authorities; even some proponents of such optimal tolls have suggested that they be returned to motorists through a lump-sum redistribution of some sort. (Keeler and Small, 1977: 23)

However, the policy suggestion of returning excess revenues to motorists may have perverse consequences as this may incentivise motorists to revert back to their previous behaviour (Emmerink, 1998:47). Baumol and Oates (1988:23-25), too, are adamant that those who bear externalities should not be compensated.

3.5. FRAMEWORK FOR ANALYSIS

Having laid out the broad terms for evaluating the economics of road use, a more rigorous framework is developed below. It is based on Newbery (1989) and sets about deriving the externalities arising from road use, with the focus on pricing congestive and road damage externalities.³⁵

This framework begins with the assumption of a constant traffic flow of N_i vehicles of type i per year. Each of these type i vehicles has an average congestive effect of P_i passenger car units (PCUs) and an average road damaging effect of E_i equivalent standard axle loads (ESALs) for each type i , where a single ESAL is equivalent to an axle load of approximately 8 157 kilograms (kg) or 80 kilonewtons (kN) tons (Gillespie *et al*, 1992).³⁶

³⁵ Other papers that combine congestion and road damage charges include Winston (1991) and Hau (1992).

³⁶ ESALs and load equivalency factors (LEFs) are interchangeable terms and this paper uses both.

Total traffic flow per year is given by:

$$N = \sum_i N_i P_i \quad [1]$$

Each vehicle operates at a cost of v_i cents per kilometre (km), given by:

$$v_i = v_{i0} + t_i + b_i R \quad [2]$$

where v_{i0} is the vehicle overhead and operating costs (depreciation, fuel, maintenance, insurance); t_i is the time cost for the occupants; and $b_i R$ the damage done to the vehicle by the road surface (measured by its average roughness R).³⁷

The time taken to travel each kilometre is dependent on the traffic flow N as well as the capacity of the road h , which is a function of its width w in the number of standard width lanes. Road capacity is implicitly defined as being homogenous of degree zero in both N and h (Newbery, 1989). In other words, the capacity of the road has doubled if it can carry twice the traffic at a given speed, as given by:

$$t_i = t_i\{N, h(w)\}$$

The annualised cost per kilometre of road construction is a function of the road's strength S and width w :³⁸

$$K = K_0 + wK(S) \quad [3]$$

where strength S is defined as being proportional to thickness and thus the marginal cost of strengthening a road is constant.³⁹ $K_0 = 0$ is for the strict case where there are constant returns to road expansion.⁴⁰

The annual maintenance cost per kilometre is given by:⁴¹

$$wM(D, S)$$

where M is the maintenance cost per lane (and so wM gives the cost for all lanes), and is a function of the annual number of ESALs experienced by the most heavily trafficked lane, D , and road strength, S . Newbery (1989:190) states that it is commonly found that the outer, or

³⁷ Newbery (1989) acknowledges that this cost is marginal and, practically, can be ignored.

³⁸ This paper interprets annualised cost as being the cost of road construction, spread out over the lifespan of the road

³⁹ Newbery (1989) defines $\frac{dK}{dS}$ as constant.

⁴⁰ There is some debate as to whether this holds in practice, as will be seen in section 3.4.

⁴¹ Not yet incurred by the road agency, but rather measures that portion of maintenance costs that are attributable to a given year.

slow, lane (on a multi-lane road) attracts the most traffic by ESAL.⁴² The fraction of time spent by heavy vehicles in the slow lane is given by

$$D = g(w) \sum_i N_i E_i$$

Where E_i is negligible for all but heavy vehicles. Newbery (1989) argues this is the correct function as it is standard practice to build the entire road to the same strength and then repair all lanes at the same time upon the first reaching a certain minimum threshold.

The total annual social cost per kilometre of road is thus:

$$F = K_0 + wK(S) + \sum_i N_i(v_{i0} + t_i + b_i R) + wM(D, S) \quad [4]$$

Or alternatively:

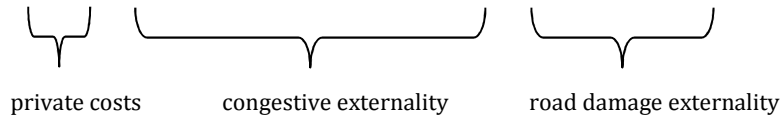
$$F = K + \sum_i N_i(v_{i0} + t_i + b_i R) + wM(g(w) \sum_i N_i E_i, S) \quad [5]$$

The marginal social cost of an extra vehicle of type i on the road is found by differentiating equation [4] with respect to N_i . Newbery (1989) derives:

$$\frac{\partial F}{\partial N_i} = v_i + \sum_j N_j \left(\frac{\partial t_j}{\partial N} \right) P_i + w \left(\frac{\partial M}{\partial D} \right) g(w) E_i \quad [6]$$

This paper, however, contends that the above does not capture the congestive effect that the marginal vehicle has on vehicles in its own group. This should therefore be amended to:

$$\frac{\partial F}{\partial N_i} = v_i + \sum_j N_j \left(\frac{\partial t_j}{\partial N} \right) P_i + (N_i - 1) \left(\frac{\partial t_i}{\partial N} \right) P_i + w \left(\frac{\partial M}{\partial D} \right) g(w) E_i \quad [7]$$



This is where $i \neq j$.

The first term is the private cost borne by the marginal individual user. The second and third terms capture the congestive externality on all other road users, while the fourth term provides the road damage externality.

⁴² Evidence from the US shows that heavy vehicles travel primarily in the slow lane on a multi-lane road, allowing faster vehicles to pass in other lanes. Because this relationship is primarily technical, it is likely to hold generally (Newbery, 1989:190).

Though not included in this derivation, the paper here proposes one further externality – that of incremental construction costs. This will be fully introduced and discussed in the next section.

There are two other broad categories of externality relevant to road use, namely environmental externalities (air and noise pollution) and accident externalities.⁴³ These externalities are not the focus of this paper and will not be dealt with in any further detail other than to say that it is unclear whether toll roads are better than untolled roads in respect of these issues. For the interested reader, discussions of these externalities can be found in Maddison *et al* (1996) and Newbery (1988b).

3.6. ON WHICH ROADS?

The existence of externalities for road use in general does not, however, necessitate that all parts of the road network are priced with Pigouvian taxes. Vickrey (1963) argued that lightly trafficked rural roads and suburban streets should be exempt of road pricing, as marginal cost pricing does not make sense given that marginal social cost falls below the average costs for these parts of the network.

Roads that should be priced are those that generally require substantial investment and suffer from chronic congestion are arterial routes and the highway system. Heavily trafficked roads in urban centres should also have a price associated with their use, particularly with regard to the congestive effect of trucks in these areas (Vickrey, 1963).

The pricing construct above is focused on the pricing of congestive externalities. The existence of the road damage externalities, however, requires pricing on all roads for heavy vehicles, the rationale and mechanics of which will be explained later in the paper.

Furthermore, the introduction of the incremental construction cost externality brings about its own complications for deciding which roads to price. This will be addressed in section 4.6.

⁴³ Newbery (1989) notes one other, the road user damage externality, where extra vehicles increase the roughness of the road, increasing the damage for other vehicles using the same road, though this externality is quantitatively negligible and can be ignored for practical purposes.

4. PARTIAL EQUILIBRIUM ANALYSIS OF ROAD PRICING

4.1. APPLYING THE THEORY

This section enquires whether current South Africa road pricing adequately prices the externalities that road users generate. Limited literature exists locally for such an exercise, with Freeman (1981) the seminal paper in this area. The paper here evaluates South African policy against the theoretical benchmark of optimal externality-corrective pricing for the three externalities sketched out at the end of the previous section, namely the congestive, road damage and incremental construction cost externalities.

To the extent possible, data has been used to estimate externality values. Where there are data constraints, the matter is conceptually discussed with some reference to the South African case.

4.2. THE CONGESTIVE EXTERNALITY

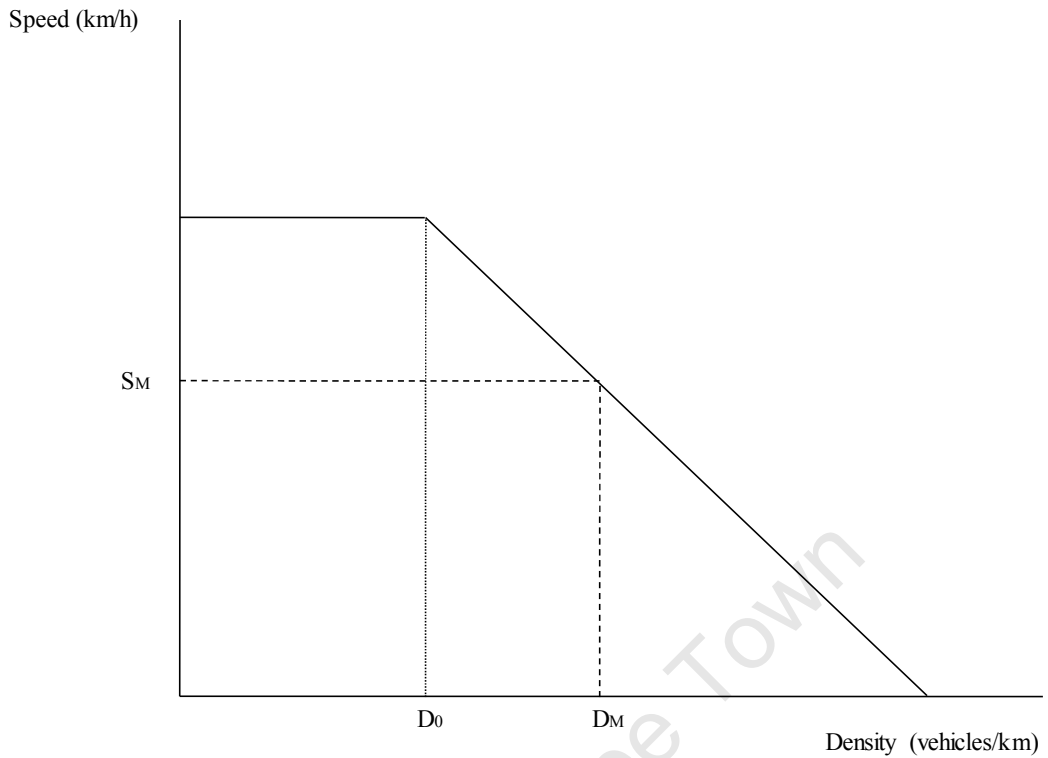
An intuitive understanding of the congestion externality is provided by Hau (1992; 1998), incorporated for the purposes of this paper. It takes the form of a graphical microeconomic framework that involves the adaption of technical transport engineering concepts into applied economics.

Consider a representative driver travelling along a given stretch of road with a fixed beginning and end point and fixed capacity, given by road width. Assume that this driver is the only vehicle, with congestive effect of P_i PCUs on the road.

Initially the entry of other vehicles will have no effect on travel time, but, *ceteris paribus*, beyond some vehicle density additional vehicles will increase the travel time for all drivers. Given fixed road capacity, traffic density increases as the number of vehicles increases. Speed decreases and therefore travel time lengthens. Traffic flow is therefore the product of traffic density and traffic speed.

The maximum flow for a road is given by (S_M, D_M) , the rectangle under which the area on the graph is greatest. Gerlough and Hubner (1975; cited in Hau, 1998) suggest that the real-world capacity is approximately 1800 vehicles per lane per hour at an average speed of 55km/h.

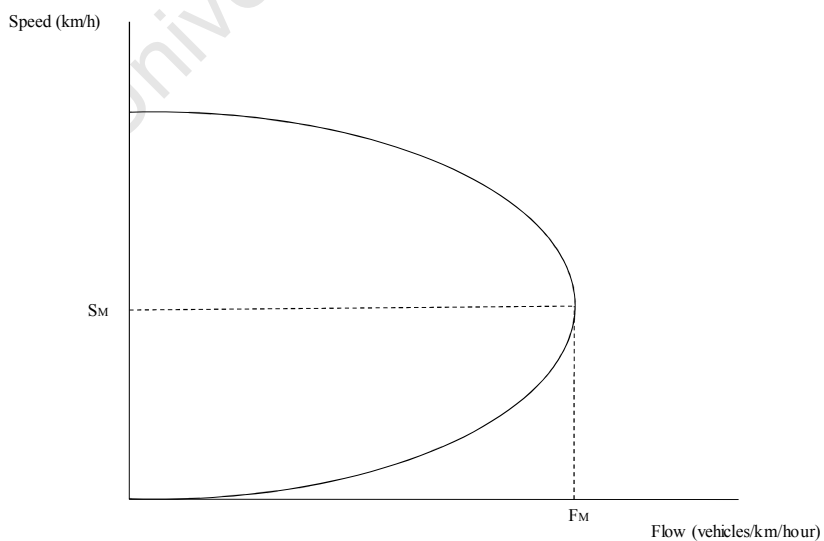
Figure 4: Speed-density curve



Source: adapted from Hau (1998)

Figure 5 below plots the relationship between speed and flow. At low speeds traffic flow is correspondingly low. As speeds increase, flow increases until a maximum of F_M with corresponding speed S_M . For speeds above S_M , travelling distance between vehicles increases (to prevent accidents) to such an extent that traffic flow decreases.

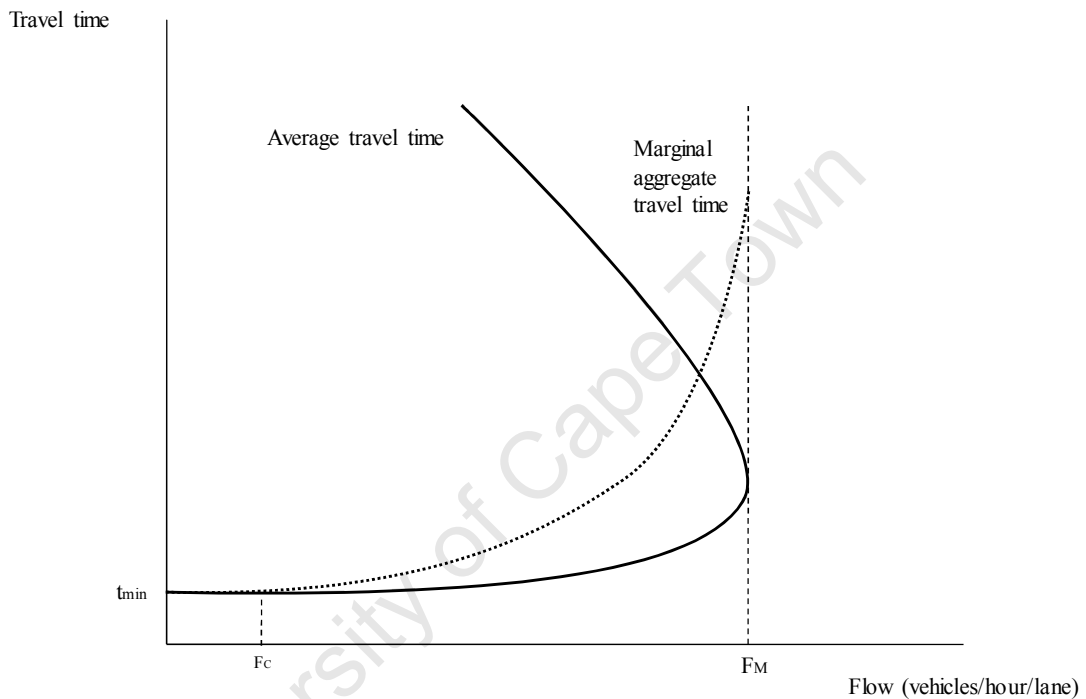
Figure 5: Traffic flow curve



Source: adapted from Hau (1998)

This speed diagram can be converted to a travel time diagram, given that the length of the road is known.⁴⁴ This is shown in Figure 6 below. As flow approaches F_M , travel time increases gradually. Upon reaching F_M , the travel time for the average user increases rapidly while traffic flow simultaneously falls. This backward bending section is the situation of congestion. Marginal aggregate travel time also rises steeply as one approaches F_M . In other words, an additional road user causes the travel time of all existing road users to increase by some fraction, and that same user also endures a prolonged travel time.

Figure 6: Travel time - flow curve



Source: adapted from Hau (1998)

These diagrams illustrate the classic congestion problem in terms of technical traffic engineering concepts. To represent the relationship as an economic one, the value of the drivers' travel time takes on the critical role of the "shadow price" of road use.⁴⁵

Time cost is not the only cost incurred when making the decision to embark on a trip using a particular road.⁴⁶ There are essentially three types of private costs incurred by the driver: vehicle fixed costs (overheads); variable costs related to vehicle use (fuel, oil, maintenance and depreciation); and time costs of the driver (and/or occupants).

⁴⁴ With distance and speed, time can be easily calculated given that $speed = \frac{distance}{time}$

⁴⁵ Importantly, the value of commuting time is less than the value of working time. This is because it is mostly leisure time that is sacrificed in traffic congestion, not working time.

⁴⁶ A peculiarity with road transport is that the driver plays the role of both a consumer and producer (Mohring, 1970). A consumer as the user demands transport services. A producer as the driver supplies most of the necessary elements for consumption (vehicle, fuel, maintenance, etc.).

The time component plus the vehicle variable costs give the average variable cost (AVC). This AVC takes on the same shape as the average travel time curve in Figure 6. It is therefore the time component that gives the AVC its upward slope toward F_M and its backward-bending slope thereafter. The AVC curve climbs upwards because of the negative interactions that occur between motorists before they reach the engineering capacity of the road, Q_{MAX} .

Variable costs of vehicle use, predominantly fuel and maintenance, do not affect the slope of the AVC as they are said to be independent of traffic flow (Mohring, 1976). Most drivers recognise that at low speeds (and therefore high traffic volumes), fuel costs are high due to the stop-start nature of congestion (Maddison *et al.*, 1996). However, it is also known that high speed (where traffic volumes are lower) tends to raise fuel consumption (Hau, 1998:46). Mohring (1970) believed that these two opposing effects roughly cancelled each other out. In other words, fuel consumption can be ignored when considering speed. Keeler and Small (1977) substantiate this, saying “the only important way in which travel costs and highway speed are likely to be related for freeway travel is through the value of travel time.” Moreover, this assumption has been found to be valid in South Africa (Pienaar and Nel, 2009:46-47).

The AVC curve mimics a supply curve in a world of congestion. This is because AVC represents the costs incurred by the representative driver. The marginal cost (MC) curve includes the external costs caused by the additional driver (the congestive effect) and therefore only becomes the supply curve when this externality is explicitly priced.

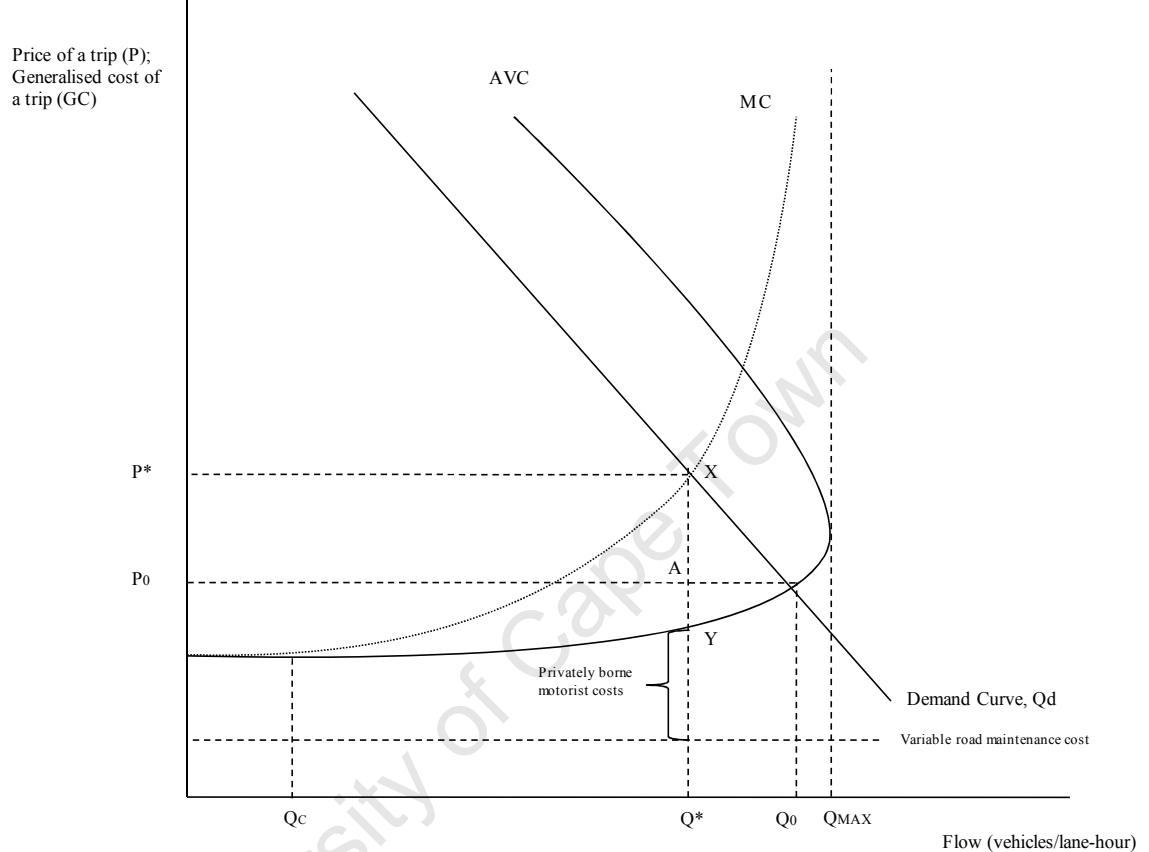
As a result, the intersection of the AVC and the demand curve Q_d at $(Q_0; P_0)$ in Figure 7 is a stable equilibrium in the non-toll scenario. It is an equilibrium point because the motorists’ willingness-to-pay curve (demand curve) intersects the supply curve, AVC (Hau, 1998:46). At this point, there is the “socially excessive consumption of road trips” (Verhoef, 2008:7).

The efficient road user charge for congestion is therefore the marginal external congestion toll (distance XY), given by the vertical difference between MC and AVC. It is the monetised deviation between marginal and aggregative travel time from Figure 6. In other words, MC contains the additional travel time increments that are attributable to the marginal user.⁴⁷ The optimal level traffic is therefore Q^* where MC intersects the demand curve at X ($Q^*; P^*$). This cost is equal to the increment in average time cost caused by the motorist entering the road multiplied by the number of vehicles in the traffic stream. Other components of the overall cost

⁴⁷ With respect to these time losses due to congestion, it is important to remember that the time wastage remains internal among the group of road users locked in traffic congestion – it is not a true externality spilling over to non-users.

are the motorist's private costs, other variable costs, and the road maintenance cost (Hau, 1998:47).⁴⁸

Figure 7: Deriving the cost curves and optimal toll



Source: adapted from Hau (1998)

The attraction of setting a congestion charge that equates the marginal cost with average variable cost is that it maximises efficiency by allowing people to respond by making a choice between the options available. Verhoef (2008:18) argues that this is biggest advantage to road pricing over other policies. A policy such as direct regulation of trips (through, for example, allowing cars with even-numbered licence plates on a road one day and odd-numbered plates the next) would achieve the same outcome but would discourage trips that have a net social benefit (Verhoef, 2008:11). In other words, road prices are useful as they act as a mechanism to nudge people towards different behaviour. According to Verhoef (2008) and Newbery (1989:173), setting an optimally priced toll, will allow road users to respond in a multitude of ways, including:

- making fewer trips;

⁴⁸ Where the road maintenance cost is incurred through general taxes it will be shared by non-users.

- adjusting their route to incorporate less congested streets;
- adjusting the time at which they travel (for example, peak congestion charges promote travel at quieter times of the day);
- substituting to different transport modes (in the case of commuters, carpooling or public transport; for transport firms, rail freight instead of road freight);
- adjusting work or residential location; or
- remaining on the road and paying the toll.

Importantly, to eliminate congestion completely, the toll would have to discourage traffic to the point Q_C in Figure 7 (and F_C in Figure 6), where the marginal cost (travel time) begins to deviate from the average. Reducing traffic to this level is not efficient as it unnecessarily restricts trips where drivers place a higher value on travel than the cost they would incur by taking the trip.

4.3. IMPLICATIONS OF THE CONGESTION TOLL

Though optimality follows neatly from the theory, there are several complications to congestion charges that should be noted.

4.3.1. Welfare impact

Though social welfare improves as a result of congestion charges, its gains are not necessarily equitably distributed. The price system generally presumes that income distribution is a non-issue, with perfect competition gives an ideal welfare outcome provided one is content with the initial distribution of income. From a welfare perspective, the wealthy may be more willing to pay to avoid congestion, but the poor may not have the option. This presents a classic welfare conundrum.

To motorists that remain on the road, the “tolled”, the congestion charge is akin to a tax increase. This is because the optimal toll XY exceeds the individual motorist’s time saving AY and the “tolled” are therefore worse off by the distance XA (Hau, 1998:47). Even when scaling up, the aggregate toll payment still exceeds the total value of time savings.⁴⁹ The “tolled off” are also worse off because they are either pushed to a public transport mode or travel at different times.

⁴⁹ Though the assumption underlying this model is motorists with homogenous time valuation, this holds under heterogeneous time valuations, too – see Hau (1992, appendix).

A further complication for congestion charges is the elasticity of demand for road use. The value transport users place on service quality is significantly higher than for other services, as a proportion of the users' wage rate (Winston, 1985). If demand is highly inelastic, then the likelihood of meaningful cuts in congestion will be reduced. Therefore, the pricing will not have the desired effect on congestion given the non-optimal pricing and will simply act as a tax.

Without redistribution (and except in the case of hyper-congestion), the main beneficiaries of road pricing are government and, under the relaxed assumption of heterogeneity in time valuation, users with a high monetary time value (Hau, 1992 and 1998). The fact that the latter group benefits under road pricing is a perverse outcome in the South African context.

Welfare effects are therefore an important consideration in the decision to introduce congestion charges. Indeed, Newbery (1983:183) argues that the greatest difficulty with road pricing is "restructuring the current system of road charges so that a sufficiently large fraction of road users are, and perceive themselves to be, better off under the new system".

4.3.2. Differing congestive effects of vehicles

Recall that each vehicle on the road has some congestive effect, P_i . Though all vehicles may cause congestion, some are more congestive than others. Due to their larger size and weight, heavy vehicles are slower and less agile on the road as compared to light passenger vehicles. Additionally, they tend to be slower moving, especially on hills and turns. Heavy vehicles therefore have a higher congestive effect as they cause traffic density to increase more rapidly than light vehicles.

In order to account for the different congestive effects of light and heavy vehicles, the passenger car unit (PCU) measure is used as an equilibrating factor for the congestive effect of a vehicle.⁵⁰ It is suggested that where 1 PCU is the congestive effect of a passenger car, a PCU measure in the order of 2 or 3 would represent a truck "depending on the size and power of the truck, the road geometry and traffic conditions" (Newbery, 1989:166). These vehicles therefore cannot be considered to warrant the same magnitude of congestion charge as a light passenger vehicle.

4.3.3. Peak congestion charges

An important implication of the model is that peak-hour use, when congestion is at its most severe, should be charged much more than off-peak travel. Vickrey argues that "[e]ven if

⁵⁰ This measure is analogous to the passenger car equivalents (PCEs) used in the US.

urban motorists on the average paid the full cost of the urban facilities, rush hour use would still be seriously underpriced” (Vickrey, 1963:455).

Excess supply of road capacity, as with most public utilities, cannot be stored (Mohring, 1970). Therefore, charging appropriate prices that vary with demand peaks and troughs for a resource with fixed supply is efficient. Vickrey found that, without tolls at peak hour, the equilibrium state would involve varying degrees of queuing, with those morning commuters “arriving at their offices closest to their desired times generally having to spend relatively more time in the queue than those who choose to push their arrival time further away from the desired time” (Vickrey, 1969:253).

SANRAL has not implemented toll price differentiation by time of day in any existing toll plazas around the country. This is probably because congestion is not a problem on most toll roads, as these roads are generally inter-city rather than urban roads.⁵¹ Under the impending GFIP, however, which is located in a high traffic density area, an attempt at differentiating toll prices by time of travel is to be made. Those drivers who travel in off-peak times face lower charges than those that travel during the rush hour commute. This will be further explored in section 6.

4.4. THE ROAD DAMAGE EXTERNALITY

Prior to Keeler and Small (1977), Newbery (1989), and Hau (1992), the standard theoretical model assumed an infinitely durable highway. Given the high level of maintenance expenditures roads require, this is clearly unrealistic (Newbery, 1989).

All vehicles do some quantum of damage to the road surface and structure when using a road. Road damage is visually observable through the phenomena of rutting (pavement deformation) and pavement fatigue, which leads to cracking (Gillespie *et al*, 1992).

There are two broad determinants of road damage. Firstly, road damage is partly dependent on the road itself and its location. This includes natural or pre-existing factors such as road surface type and weather conditions. Although Paterson (1986; cited in Newbery, 1989) found that almost 60 percent of road damage in the UK was due to weather conditions, this was primarily caused by “frost heaves” that occur when a road is under snow or ice, and are therefore not relevant in South Africa given the country’s temperate climate. However, it should be noted that the dryness of the road affects the level of damage a vehicle imposes on the road, with wetter roads more susceptible to damage. General factors that also matter in road design are

⁵¹ Volumes carried on the N3 may, however, necessitate congestion charges at times, especially over long weekends and the festive season.

pavement composition, material types, definition of terminal conditions and “rideability” (CSIR, 2000).

Secondly, vehicle-specific factors related to the design of the individual vehicle affect road damage. Factors particularly important in determining the impact a particular vehicle has on the road are the vehicle mass and axle configuration. Jointly, these determine axle load, the mass exerted on the road by each axle of the vehicle. Other relevant vehicle-specific factors include tyre pressure, tyre configuration, suspension type and axle configuration (Gillespie *et al*, 1992).

Recall that road damaging power is measured by cumulative ESALs, also known as load equivalency factors (LEFs). The relationship between axle load and road damage is an exponential one, originally empirically measured by the American Association of State Highway and Transportation Officials (AAHSTO) in the late 1950s. Importantly, the damaging power of a vehicle is proportional to the fourth power of the load on an axle, commonly referred to as the “fourth-power law” (Hau, 1992:48).

The formula for calculating approximate LEFs is given by:

$$LEF = \left(\frac{P}{80kN} \right)^n \quad [8]$$

where n is 4.2 (i.e. the exponent in the fourth-power law) and P is the axle load of the vehicle in kN (CSIR, 2000). The LEF factor equates the road damage of a particular vehicle relative to the damage done by a base comparison vehicle (80kN/8.16 ton axle load with 520kPa tyre pressure).⁵²

The load equivalency for various vehicle axle-loads is given in Table 7 below, produced from equation 8.

This table shows the exponential impact of axle load on road damage. The LEF of a dual-axle light passenger vehicle, weighing 1 500kg, is approximately 0.0001. A 49-ton truck with seven axles, on the other hand, has a total vehicle LEF of 3.68. For 10 000 passenger vehicles on the road, the comparable road damaging factor would equate to an LEF of 1.0, substantially below the damage from the single truck.

Should the aforesaid 49-ton truck choose to overload with, say, 10 tons on each axle (a gross vehicle mass of 70 tons), the single-axle LEF would be 2.35 and the total vehicle LEF 16.46.

⁵² This is the standard vehicle type for comparison by LEF/ESAL, providing a value of 1.0.

Over 185 000 passenger vehicles would have to pass over the road before the damage between the two is equivalent.

Table 7: Single-axle equivalency factors

Single axle load (kN)	Axle mass range (kg)		LEF / ESAL
<15	0	1530	0
15 - 24	1530	2447	0.004
25 - 34	2549	3467	0.019
35 - 44	3569	4487	0.062
45 - 54	4589	5506	0.15
55 - 64	5608	6526	0.32
65 - 74	6628	7546	0.59
75 - 84	7648	8566	1.0
85 - 94	8668	9585	1.6
95 - 104	9687	10605	2.4
105 - 114	10707	11625	3.6
115 - 124	11727	12644	5.1
125 - 134	12746	13664	7.0
135 - 144	13766	14684	9.4
145 - 154	14786	15704	12
155 - 164	15806	16723	16
165 - 174	16825	17743	20
175 - 184	17845	18763	26
185 - 194	18865	19782	32
195 - 204	19884	20802	39
>205	20904	-	≥50

Source: adapted from CSIR (2000:ch.8, 12)

This relationship implies that road damage is predominantly generated by heavy vehicles such as trucks and buses, while passenger cars have a negligible effect on the road surface (Newbery, 1989:166). This is a striking fact. In charging road users for the damage they do, light passenger vehicles would accrue almost no charges while heavy vehicles would be left to fund nearly all road maintenance.

4.5. QUANTIFYING THE ROAD DAMAGE EXTERNALITY

In South Africa, the issue of road damage is particularly important given that the majority of general freight in the country is transported using trucks.⁵³ While it is axle load that ultimately determines road damage, an increase in GVM results in higher road damage with reduced relative cost recovery, all other things equal. The legal gross vehicle mass (GVM) limit for

⁵³ General freight excludes bulk mining products, such as iron ore, coal and manganese.

trucks in South Africa was increased from 48 000kg in the early 1990s to 56 000kg (Leiman, 2003:253).

The table below compares the GVM and gross combination mass (GCM) limits across various countries including South Africa.⁵⁴ While it is acknowledged that road regulations may specify restrictions based on axle load, tyre load, *etc.*, these mass limit measures provide a high-level assessment of various countries' regulatory approaches toward vehicle loadings.

From the table it is evident that South Africa's mass limit is uniformly higher, by a minimum of 27 percent. All other things equal, heavy trucks in South Africa are legally permitted to do more damage to the road than compared to Australia, Germany, New Zealand, the United Kingdom and California.

Table 8: Comparison table of legal maximum gross vehicle masses (GVM) in 2012

Country	Vehicle type (rigid or combination)	Gross legal mass (kg)
South Africa	Either	56 000
New Zealand	Rigid	44 000
United Kingdom	Either	44 000
Germany	Combination	44 000
Australia	Either	42 500
New Zealand	Combination	39 000
California, United States	Either	36 287

Source: Various road traffic authorities' publications

Note 1: The legal mass limits given here are for vehicles with unrestricted access to road network.

This difference is not necessarily problematic, however, provided that the recovery of the corresponding road damage is priced accordingly. What therefore follows is a quantification of road damage done by heavy vehicles in South Africa and the relative cost recovery by authorities of this damage. In doing so, it attempts to gauge the current pricing position and establish whether there is an under- or over-recovery of this damage.

The road damage cost-recovery model consists of two parts. The first estimates the damage inflicted on South African roads by a set of representative heavy vehicles, while the second approximates the recovery of this damage by authorities.

Both of these quantifications incorporate a variety of technical specifications and assumptions. The output of the model is a set of distance-adjusted damage costs and user charges for the representative vehicles. For comparability, all figures are provided on a per-kilometre basis.

⁵⁴ Gross combination mass is the total weight a combination vehicle can carry and tow. Gross vehicle mass applies to rigid vehicles (i.e. those that are not fitted with trailers).

4.5.1. Road damage

Representative heavy vehicles and their specifications are drawn from a technical CSIR report, in which were selected given their “popularity”, particularly as carriers on long-haul routes (Roux and Nordengen, 2010:16). Information is provided for ten representative freight trucks, the details of which are given in the table below.

Table 9: Vehicle specifications of representative road freight trucks

Details	Tare Mass (kg)	Gross Mass (kg)	Payload (kg)	Number of Tyres	Number of axles
Truck 1	10 700	26 000	15 300	10	3
Truck 2	17 400	42 000	24 600	18	5
Truck 2a	17 400	42 000	24 600	12	5
Truck 3	17 200	44 000	26 800	18	5
Truck 4	18 000	50 000	32 000	22	6
Truck 4a	18 000	50 000	32 000	16	6
Truck 5	27 900	54 000	26 100	22	6
Truck 6	20 500	56 000	35 500	26	7
Truck 7	22 100	56 000	33 900	26	7
Truck 8	23 200	56 000	32 800	30	8
Truck 8a	23 200	56 000	32 800	24	8

Source: Roux and Nordengen (2010); author’s own calculations

Note: Payload has been inferred from the gross vehicle mass and tare mass.

For the purposes of this model, LEFs are based on the South African Mechanistic-Empirical Design Method (SAMDM) calculations, obtained from Roux and Nordengen (2010). The LEFs generated are for a variety of pavement types in South Africa, in both wet and dry conditions. For the purposes of this paper, an LEF for each vehicle is obtained by averaging across the five pavement types used in the study and adjusting the wet and dry measurements specific to South Africa’s climate.⁵⁵

The damage factor for each vehicle is then converted into a monetary value by multiplying the LEFs by the “mass fee”. This fee is determined by the DoT on an annual basis, and is based on the cost of constructing and maintaining a one kilometre stretch of single-lane road.⁵⁶ For 2012, the mass fee was R0.59 per km (Department of Transport, 2012a).⁵⁷

Table 10 gives the climate-adjusted LEFs and associated damage values per vehicle.

⁵⁵ South Africa experiences approximately 101 days of rainfall per year and therefore wet conditions account for approximately 27.7 percent of the days road travel occurs on. See <<http://www.south-africa.climateemps.com>>. This assumption is approximate only, as it generally does not rain for a full day.

⁵⁶ This mass fee is used to calculate permit fees for “abnormal load” vehicles, a point which will be touched on later in the paper.

⁵⁷ The technical relationship that stipulates an entire road must be resurfaced upon reaching a terminal roughness (Newbery, 1989) suggests the DoT’s mass fee, based on the cost for a single lane only, is underpriced.

Table 10: Road damage for representative vehicles

Details	Climate-adjusted average	Climate-adjusted road damage	
	LEF		per vehicle-km
Truck 1	3.60	R	2.13
Truck 2	4.79	R	2.83
Truck 2a	9.89	R	5.84
Truck 3	5.12	R	3.02
Truck 4	5.09	R	3.00
Truck 4a	10.17	R	6.00
Truck 5	6.31	R	3.72
Truck 6	4.35	R	2.57
Truck 7	4.48	R	2.65
Truck 8	3.42	R	2.02
Truck 8a	6.39	R	3.77
<i>Mean</i>	5.78	<i>R</i>	3.41
<i>Median</i>	5.09	<i>R</i>	3.00

Source: Author's own calculations; Roux and Nordengen (2010)

4.5.2. Cost recovery

The second line of development in the model is the calculation of a per-kilometre user charge, which requires revisiting the notion of a quasi-market price based on a basket of price interventions (licence fees; fuel levy; tolls).⁵⁸

Licence fees

Normal heavy commercial vehicles (like rigid trucks and combination trucks and trailers) pay licence fees based on gross unladen mass.⁵⁹ Fee schedules are published by each provincial traffic authority on an annual basis.

The licence fee is directly proportional to the weight of the truck and/ or its trailer, which suggests that authorities may see the fees as a proxy for road damage, albeit in the form of a lump-sum tax. Licence fees recover the administrative expenses of the licensing authority, though are also commonly applied in part to recover the fixed costs of roads (Pienaar, 2005:14).⁶⁰ An absence of appropriate data on this apportionment means that the conservative assumption is made here that the full amount collected is used for road maintenance and can therefore be attributed to recovery of road damage.

⁵⁸ Congestion charges are not included in this quasi market price as none are currently charged in South Africa.

⁵⁹ Under regulation 21(1)(g) of the National Road Traffic Act 93 of 1996, abnormal vehicles may apply for special classification and receive reduced licence fees. However, it will be shown that these vehicles are made to purchase permits that attempt to recoup road damage.

⁶⁰ In addition to paying conventional licence fees, freight carriers wishing to convey goods to and from beyond the country's borders must apply for a permit to conduct such business. These permit fees are generally not regarded as a form of road cost recovery and also have no bearing on the characteristics of the vehicle (Pienaar, 2005:15).

Table 11: Licence fee structure for commercial vehicles

Province	TRUCK					TRAILOR				
	Base fee		Incremental fee above 12 000kg			Base fee		Incremental fee above 12 000kg		
	(10 501-11 000kg)	(11 501-12 000kg)	(per 500kg or part thereof)			(up to 12 000kg)	(per 500kg or part thereof)			
Free State	R	12 012	R	14 856	R	2 048	R	14 452	R	2 048
Gauteng	R	15 132	R	18 024	R	1 224	R	18 108	R	1 476
KZN	R	17 838	R	21 072	R	1 893	R	21 072	R	1 893
Limpopo	R	11 567	R	13 690	R	1 172	R	10 890	R	984
Mpumalanga	R	10 080	R	11 928	R	1 020	R	11 928	R	1 020
Eastern Cape	R	15 468	R	18 312	R	1 632	R	19 830	R	1 632
North-West	R	13 110	R	15 498	R	1 518	R	15 450	R	1 500
Northern Cape	R	12 810	R	15 132	R	1 506	R	15 132	R	1 506
Western Cape	R	14 037	R	16 551	R	1 743	R	16 551	R	1 743

Source: Various gazetted provincial licence fee schedules

Note 1: All weights are tare (unladen) weights.

Note 2: Kwa-Zulu Natal fees are for 2011; Limpopo truck fee are for 2011 and trailer fees 2004; Western Cape fees are from 2005. This does not necessarily mean these fee schedules are out of date. It may also be the case that the provincial authorities have chosen not to increase fees in the intervening period.

As a proxy for road damage, there are several shortcomings with the licence fee measure. Firstly, the relationship between mass and damage is not linear but exponential. Secondly, the licence fees are charged based on unladen rather than laden mass, which therefore underestimates road damaging power. Thirdly, the true determinant of road damage is axle mass rather than gross vehicle mass, which the regulations do not take into account.

Fuel levy

The levy on diesel, the predominant fuel type for heavy vehicles, is 183c per litre for the 2012/2013 fiscal year (Department of Energy, 2012). This charge is also more or less linearly proportional to road use and mass, rather than exponentially related to axle mass.

Tolls

To fully account for cost recovery, average toll costs for two of the busiest long-haul freight routes in the country's road network (the N1 from Johannesburg to Cape Town and N3 from Johannesburg to Durban) have been taken into account by making the assumption that the representative trucks travel on only these two routes for the entire assumed distance. This should therefore be seen as a conservative assumption that errs on the side of over-recovery.

4.5.3. Additional assumptions and manipulations required

There are several additional assumptions and manipulations of the data that must be made to produce readily interpretable results.

Average fuel consumption. Average fuel consumption for the representative vehicles is taken from an International Energy Agency (IEA) report on fuel economy. It is assumed that all vehicles will have the average fuel economy of a 2010 model heavy truck/bus, estimated to be 35.9l/100km (IEA, 2012:14).

Distance travelled per annum. In line with vehicle manufacturer warranties and service plans,⁶¹ the model assumes an annual distance travelled for each vehicle of 150 000km.⁶² This is then converted into an effective full payload kilometre measure. The assumption is made that, on average, every 3 out of 4 trips are with a full load, and the fourth load is empty. In other words, the load utilisation is 75 percent and effective full-load distance is therefore 112 500km.⁶³

Calculating the average licence fee. The National Traffic Information System (eNaTIS) does not provide any information to the public on the specific vehicle types registered in South Africa. By broad vehicle category, the December 2012 statistics indicated that there were 342 131 heavy-load trucks (with mass greater than 3500kg) and 164 908 heavy-load trailers registered with authorities.

Table 12: Truck and trailer population breakdown (December 2012)

Vehicle class	FS	GP	KZN	LP	MP	EC	NW	NC	WC	Total
Heavy load trucks (GVM >3500kg)	20 116	130 302	50 941	21 889	32 522	22 609	17 450	9 367	36 935	342 131
% of trucks in province	5.9%	38.1%	14.9%	6.4%	9.5%	6.6%	5.1%	2.7%	10.8%	100.0%
Heavy load trailers (GVM >3500kg)	15 158	55 301	26 569	7 107	21 240	8 150	10 627	5 158	15 598	164 908
% of trailers in province	9.2%	33.5%	16.1%	4.3%	12.9%	4.9%	6.4%	3.1%	9.5%	100.0%
Total trucks and trailers	35 274	185 603	77 510	28 996	53 762	30 759	28 077	14 525	52 533	507 039
% of total in province	7.0%	36.6%	15.3%	5.7%	10.6%	6.1%	5.5%	2.9%	10.4%	100.0%

Source: eNaTIS (2012) Live Vehicle Population (31 December 2012)

Assuming that the representative vehicles in the model are distributed across the provinces according to the proportions in Table 12, a weighted licence fee can be calculated.

Table 13: Weighted average licence fees for representative vehicles

Details	Tare Mass (kg)	Gross Mass (kg)	Payload (kg)	Weighted average licence fee
Truck 1	10 700	26 000	15 300	R 14 318
Truck 2 and 2a	17 400	42 000	24 600	R 24 506
Truck 3	17 200	44 000	26 800	R 24 506
Truck 4 and 4a	18 000	50 000	32 000	R 25 561
Truck 5	27 900	54 000	26 100	R 46 663
Truck 6	20 500	56 000	35 500	R 30 836
Truck 7	22 100	56 000	33 900	R 35 057
Truck 8 and 8a	23 200	56 000	32 800	R 37 167

Source: Author's own calculations

⁶¹ Usually 3 year/450 000km plans. See, for instance, < http://www.mercedes-benz.co.za/content/south_africa/mpc/mpc_south_africa_website/en/home_mpc/truck_home/home/trucks/actros/rep_actros_warranty.html>

⁶² This appears a reasonable assumption, especially for commercial freight vehicles involved in long haul. It represents approximately one return trip between Johannesburg and Cape Town each week of the year; for a commercial freightliner, this certainly does not appear an inordinate amount.

⁶³ Note that this assumed distance travelled per annum only affects the licence fee and the sensitivity of the overall results to this assumption is not significant.

4.5.4. Results and discussion

The tables below outline the main results of the model. For comparability, all user charges are converted to a per-kilometre measure.⁶⁴ For each of the representative vehicles, an absolute road damage value and associated cost recovery values are shown in Table 14.

Table 14: Absolute cost recovery for freight trucks

Details	Road damage	Licence fee (1)	Fuel levy (2)	Toll charge (3)	Total charge (1) + (2) + (3)
Truck 1	R 2.13	R 0.11	R 0.66	R 0.43	R 1.20
Truck 2	R 2.83	R 0.19	R 0.66	R 0.61	R 1.46
Truck 2a	R 5.84	R 0.19	R 0.66	R 0.61	R 1.46
Truck 3	R 3.02	R 0.19	R 0.66	R 0.61	R 1.46
Truck 4	R 3.00	R 0.20	R 0.66	R 0.61	R 1.47
Truck 4a	R 6.00	R 0.20	R 0.66	R 0.61	R 1.47
Truck 5	R 3.72	R 0.37	R 0.66	R 0.61	R 1.63
Truck 6	R 2.57	R 0.24	R 0.66	R 0.61	R 1.51
Truck 7	R 2.65	R 0.27	R 0.66	R 0.61	R 1.54
Truck 8	R 2.02	R 0.29	R 0.66	R 0.61	R 1.56
Truck 8a	R 3.77	R 0.29	R 0.66	R 0.61	R 1.56
Mean	R 3.41	R 0.23	R 0.66	R 0.59	R 1.48
Median	R 3.00	R 0.20	R 0.66	R 0.61	R 1.47

Source: Author's own calculations

Note: SANRAL toll rates differ for vehicles with less than 5 axles compared to those with more than 5 axles.

It is immediately evident that licence fees provide only a small fraction of the total user charge. The bulk is provided by the fuel levy and, on the constructed assumption that a vehicle travels only on toll roads, charges corresponding to the use of toll roads.

Table 15 outlines the proportional cost recovery from users of road damage done. Three graduating scenarios are provided, with the extreme right-hand column providing total cost recovery as a percentage of road damage inflicted.

The results are striking. Even under conservative assumptions, it is clear that road trucks in a *status quo* situation are dramatically underpriced in terms of the damage they impart on the road.

Considering the licence fee only, no truck pays for more than 18 percent of its road damage. On average, trucks pay less than one-tenth of the damage they impart on the road. If one considers the fuel levy as a method of cost recovery too, then cost recovery improves somewhat, to just less than one-third on average. The median truck that travels on untolled national roads and provincial and municipal roads therefore pays slightly more than one-third of the damage cost.

⁶⁴ Given fuel consumption, the fuel levy is easily converted into an effective per-kilometre charge. To convert the lump-sum licence fee per-kilometre measure, the total fee cost by the annual distance travelled. A similar conversion is done for toll charges, though on a per-trip basis.

Table 15: Proportional cost recovery for freight trucks

Details	Cost recovery - licence fee only	Cost recovery - licence fee and fuel	Cost recovery - licence fee, fuel
		levy	levy and tolls
Truck 1	5.3%	36.2%	56.4%
Truck 2	6.8%	30.0%	51.6%
Truck 2a	3.3%	14.5%	25.0%
Truck 3	6.3%	28.1%	48.3%
Truck 4	6.7%	28.5%	48.9%
Truck 4a	3.3%	14.3%	24.5%
Truck 5	9.8%	27.5%	43.9%
Truck 6	9.4%	35.0%	58.8%
Truck 7	10.4%	35.2%	58.3%
Truck 8	14.4%	47.0%	77.2%
Truck 8a	7.7%	25.1%	41.3%
Mean	6.8%	19.3%	17.4%
Median	14.4%	35.7%	48.9%

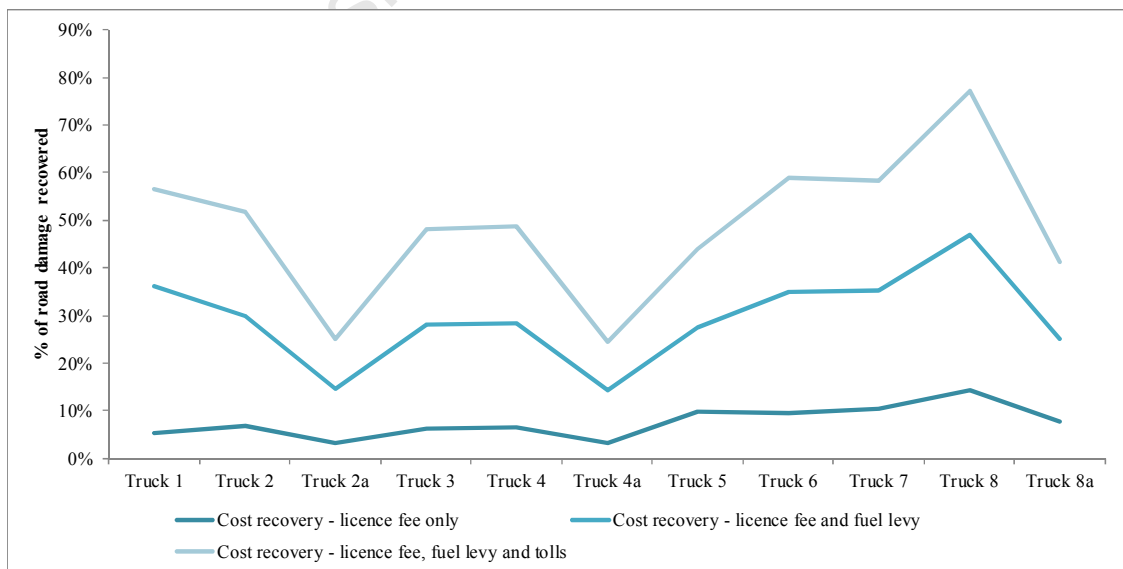
Source: Author's own calculations

Note: Mean values calculated using the average of the sum, not the average of the average.

Those sections of road that are tolled on the user pays principle show that truck users are not, in fact, adequately paying for road use. Including licence fees, fuel levies and tolls, the median truck on the two main road freight corridors will pay for just over half of the damage caused.

While it is true that the performance of trucks differs within the representative sample, only one truck ("Truck 8") approaches full cost recovery. This is despite the fact that relatively conservative assumptions were employed, which should lean towards over-estimating cost recovery.

Figure 8: Graphical representation of road damage cost recovery



Source: Author's own diagram

Taking a step back, the findings of this model suggest that there is a perverse treatment of road freight trucks *vis-à-vis* the damage these vehicles do to the road. Essentially an arrangement exists whereby taxpayers and light vehicle users are subsidising the damage done by road-freight trucks.

Given the heavy vehicle population size (approximately 342 000 registered trucks weighing greater than 3.5 tons), it is reasonable to infer that trucks are responsible for a large amount of unpriced road damage. As a substantial amount is unpriced, this cost is therefore incurred by general taxpayers along non-toll roads and users of light vehicles on toll roads.⁶⁵ Moreover, this is the calculation of only one externality charge while it incorporates almost all charges that a freight truck may incur in respect of road access and use. Such a situation indicates a severe underpricing of heavy vehicles.

4.6. THE CONSTRUCTION COST EXTERNALITY

4.6.1. A less obvious externality

An extension that this paper posits is for the pricing and recovery of those additional costs incurred in the construction of a new road that are required in order to accommodate certain types of vehicle.

Given that roads tend to cater for heterogeneous vehicle types and thereby support a combination of freight and passenger trips, the roads agency may be seen as a multiproduct firm rather than the producer of a single output (Winston, 1985). The agency therefore incurs some base level of costs to provide basic road infrastructure and thereafter incremental costs to support the specifications of various vehicle types. Regardless of the existence and direction of economies of scope, the concept of incremental cost pricing for multiproduct firms is well supported in economic theory, particularly in the field of competition economics.⁶⁶

4.6.2. Engineering roads for heavy vehicles

The capital cost of roads encompasses numerous elements. Chief amongst these are land acquisition costs and construction costs. When it comes to catering for multiple road user type, both of these cost elements are affected.

⁶⁵ The homogeneity of the truck population makes it difficult to quantify an estimate of the cumulative damage. More granularity in the vehicle population data from eNaTIS and estimates of average distances travelled would allow for such calculations.

⁶⁶ Incremental cost price benchmarks in competition economics include average avoidable cost (AAC) and long-run average incremental cost (LRAIC).

Land costs, for example, vary with the width of the road, which may be a function of the type of vehicle it is designed to support. Large vehicles like trucks and buses require more road space, so their presence on a road requires a wider road.

Construction costs vary according to the specific “level of service” the road is designed for. This level of service may be expressed as the combination of the pavement and stormwater-drainage elements of a road (CSIR, 2000). Road designers therefore have to cater for different traffic types in their pavement design. Pavement design and construction costs can be broken into two types. Firstly, those which are fixed because they must be incurred regardless of road type, such as basic earthworks and road markings and signage. Secondly, there are construction costs that vary with design requirements, such as curvature (angle of corners), gradient (road slope), bridge construction, and thickness of the road surface. A road built only for light vehicles may be relatively inexpensive, but one engineered to support heavy vehicles accentuates these expenses.

The roads agency therefore has the objective function of choosing road specifications that will be the most cost-effective given the intended design life of the road and the vehicle type distribution of its users.

4.6.3. The economic case for incremental cost recovery

What is currently missing from the road pricing discourse is the pricing of particular vehicle types whose presence on the road necessitates the construction of more expensive road infrastructure. More formally, particular vehicle classes that necessitate additional engineering and design specifications (and therefore costs) to accommodate their presence on the road should incur correspondingly higher charges for use.

Particularly, it is heavy vehicles that necessitate these additional costs, given their mass, dimensions and power-to-weight ratios. This paper suggests that there are broadly four categories of incremental costs that are relevant for these vehicles.⁶⁷

Incremental cost of general road design. Increased land costs are required for vehicles that require wider lanes and therefore also wider road reserves. Wider roads also increase the general cost of earthworks and material (gravel, tarmac, stone, etc.). Additionally, land costs may differ where routes are chosen for gentler inclines that support the travel of heavy vehicles.

⁶⁷ Although some of these incremental costs reduce road damage and congestive effect, this does not mitigate the price that users should face for incremental construction costs. Importantly, the benefits will be reflected in these other externality values.

Incremental cost of thicker road surfaces. Heavy vehicles also affect construction costs of roads through road strength. From Newbery (1989), recall that the strength of a road is proportional to its thickness and a thicker road means increased construction costs. As indicated, South Africa allows heavier trucks on its roads than most countries do. This necessitates that stronger roads be constructed.

Incremental cost of a road over hills. Construction costs are higher in the presence of heavy vehicles, as they generally require less steep inclines. Additionally, crawler lanes are constructed on national roads to alleviate the congestive effect of these vehicles on other motorists.

Incremental cost of a road over rivers and gorges. Engineering specifications are more stringent for bridges that traverse rivers and gorges in order to deal with the increased stresses imposed by heavy vehicles.

In all of these instances, the costs would otherwise not be incurred and should not be borne by other users. However, it is taxpayers and general road users who provide for the presence of heavy vehicles by funding the costly infrastructure that these vehicles necessitate. Where the general public funds these additional construction costs, an externality is clearly identifiable.

Indeed, the only category of incremental construction costs that should not be priced to heavy vehicles is those of thicker road surfaces, provided that these costs are recovered through pricing road damage externality.

In the context of roads, shared costs (such as lighting, line markings, *etc.*) could be collectively funded by all users, should infrastructure cost recovery be desirable. In other words, some base infrastructure charge for all users then, with relevant vehicle classes incurring some higher charge for the incremental costs incurred for their presence on the road.

An important consideration for the pricing of incremental construction costs is the time length over which to recover these costs. If one were to select the design life of the road (approximately 20 to 30 years), then one confronts the problem that some design and construction costs are more durable than this, such as the decision over the gradient of the road. By selecting the design life as the recovery period, one is frontloading the cost of road use and potentially overly restricting heavy vehicles in the short run. While this concern has validity, an infinite recovery horizon would provide problems of its own, as marginal recovery costs would become negligible. An alternative way of thinking about how to price may be to recover incremental costs over the financing period of the road.

The question may also arise as to whether light passenger vehicles should be culpable for incremental construction costs such as additional lanes, which they would generally be responsible for necessitating. This begins to overlap with congestion tolls, where vehicles are essentially charged for the space that they take up on the road. Further, it is necessary to evaluate what these extra lanes would cost if built for cars only.

4.7. CONSIDERING SUNK COSTS

Certain readers may enquire why pricing for either road damage or incremental construction costs matter when the road infrastructure has already been built. The concept of sunk costs is well known in economics and can be accounted for in the understanding of pricing of both these externalities.

This paper posits that the incremental construction cost externality should be priced on newly constructed and upgraded roads only. Given that even sunk costs become escapable over time, one should therefore consider the cost of additional infrastructure investments such as adding to, or upgrading certain sections of, the network

With regard to the road damage externality, when one talks of a road with a 20-year lifespan this design life is with x million ESALs passing over it. Natural factors like the weather contribute to deterioration without any traffic. It is therefore true that, absent traffic, the road surface would deteriorate to some terminal roughness and require resurfacing regardless of traffic volumes. However, this is irrelevant when pricing road damage as the authority is pricing for actual damage done with a view to renewing the road rather than the road being allowed to fall into disrepair and be abandoned. In other words, road users would only be charged for road damage if they made use of the infrastructure. This is a further advantage of a marginal charge rather than a lump-sum instrument.

4.8. CHARGING FOR INCREMENTAL CONSTRUCTION COSTS

Estimating the price of the incremental construction cost externality is not an easy task. There is a dearth of data on freight logistics statistics in general (Department of Transport, 2011b). For newly constructed roads, there is the added complexity of forming assumptions on uncertain traffic volumes and deciding on the time period over which the costs should be recovered.

Data constraints for both construction costs and traffic volumes preclude this paper from estimating incremental construction cost charges for South African roads. However, it notes

that this would be a worthwhile endeavour for future research, especially since it is not apparent that this has been adequately identified and addressed in the literature.

4.9. RELATIVE MISPRICING OF ROAD USE IN SOUTH AFRICA

From the preceding subsections, it is evident that heavy trucks generate substantially larger external costs than light passenger vehicles. Although these costs are not always easily measured, the conceptual arguments have been identified where necessary.

Given the greater externality generating power of heavy vehicles (in terms of congestion, road damage and road construction costs), one would expect *a priori* that an externality-corrective toll pricing regime would see these heavy vehicles paying a substantially higher multiple of charges than light passenger vehicles for road damage imparted on the road, the LEF of a light passenger vehicle weighing 1 500kg was calculated earlier to be approximately 0.0001, while it averaged 5.78 for heavy vehicles in the representative sample (shown in Table 10). The truck with the lowest LEF in this sample (truck 8, with an LEF of 3.42) therefore has a damage factor 38 577 times greater than the passenger vehicle.

Toll data does show that heavy vehicles pay more than light vehicles at all toll plazas in South Africa. However, the differentiation that exists is not of the magnitude expected by the analysis. While heavy vehicles with five axles or more pay multiples of between 3.8 and 5.2 over light passenger vehicle fees, this is far below what is suggested by efficient and equitable pricing.⁶⁸

Table 16: Toll factors between 5 or more axle vehicles and light passenger vehicles

Route	2008	2009	2010	2011	2012
N4	4.2	4.2	4.2	4.2	4.2
N17	4.0	4.1	4.1	4.0	4.5
N1	3.9	3.9	3.9	4.0	4.0
N2	4.4	4.4	5.2	5.2	5.2
N3	3.8	3.8	3.8	3.8	3.8

Source: Author's own calculations, based on actual SANRAL toll charges for each of the years

⁶⁸ Chapman's Peak Drive, whose tariffs are determined by the Western Cape provincial government, has a ratio of just over 4 for *medium* heavy vehicles. Heavy trucks under our taxonomy are prohibited from using the pass.

Factoring in licence fees and the fuel levy indicates that this ratio essentially remains the same.

Table 17: Per-kilometre quasi price basket ratios for heavy and light vehicles

Details	Licence fee (1)		Fuel levy (2)		Toll charge (3)		Licence fee and fuel levy (1) + (2)		Total user charge (1) + (2) + (3)	
	Truck (average)	R	0.23	R	0.66	R	0.59	R	0.89	R
Passenger car	R	0.02	R	0.18	R	0.12	R	0.20	R	0.33
<i>Factor</i>		<i>11.6</i>		<i>3.6</i>		<i>4.9</i>		<i>4.4</i>		<i>4.6</i>

Source: Author's own calculations

Note 1: Passenger vehicle licence fee assumed to be R300 per annum, on average

Note 2: Fuel consumption for passenger vehicles assumed to be 10l/100km; the fuel levy on petrol is R1.98 per litre.

Note 4: Assumed annual distance of 15 000km.

Given the disproportionate damage done by heavy vehicles, and the incremental construction costs their existence requires, it is clear that there is a severe mispricing between road users on South Africa's roads, including toll roads. The existence of this phenomenon points sharply towards the cross-subsidisation of road carriers by passenger vehicles and the taxpayer.

Evaluating road user cost recovery in South Africa in the early 1980s, Freeman concluded along similar lines, stating that "[i]n terms of current road user payments light vehicles are subsidising heavy vehicles whose cost responsibility in terms of road repair and road construction costs is not being met by the levies and fees collected" (1981:332). Worryingly, the persistence of this mispricing and associated cross-subsidisation suggests that little progress has been made in the intervening decades, despite policy pronouncements by the DoT to the contrary.

Importantly, the fact that heavy vehicles are mispriced on toll roads is not to say that they must pay thousands of times more than a light passenger vehicle at current prices. Instead, it suggests that correctly pricing the externalities of heavy vehicles should see a reduction in tolls passenger vehicle users face and an increased burden on heavy vehicles.

5. GENERAL EQUILIBRIUM CONSIDERATIONS

5.1. FROM FIRST-BEST TO SECOND-BEST PRICES

The theoretical framework investigated in the previous section argues that optimal user tolls, where price equals marginal social cost, lead to economic efficiency (Leiman, 2003).⁶⁹ The logic of this is located in a partial equilibrium setting, where the assumption is that short-run marginal cost pricing deals with the last remaining distortion in an otherwise perfectly competitive economy (Verhoef, 2008:23).

This condition, however, is generally not met. The world in which tolls are implemented is neither perfectly competitive nor free of other distortions. Road pricing can induce indirect welfare effects in markets associated with transport and tolling may therefore, in fact, increase distortions rather than decrease them (Leiman, 2003).

The fact that the partial equilibrium conditions are not met does not mean that the theory wholly dissipates its usefulness as an analytical framework. Instead, it necessitates that first-best tolls be adjusted to account for indirect welfare effects in other distorted markets (Verhoef, 2008:23), and that transport be evaluated in a general equilibrium setting. Prices that are so designed are termed “second best” in the literature, originally generalised by Lipsey and Lancaster (1956). These second-best prices, in accounting for indirect welfare effects, imply upward or downward revisions to the first-best prices (Verhoef, 2008:12). In other words, given that it is only under universal perfect competition that the partial and general equilibrium solutions coincide, a revision of prices may be necessary.

With this in mind, the analysis in this section accounts for “broad economic efficiency” or “economy-wide welfare” (Verhoef, 2008:12). Considering these wider effects provides insight into the desirability of tolling and may improve the welfare outcome upon the decision to toll (Verhoef, 2008).

⁶⁹ This includes not only congestion, road damage and incremental construction costs but also externalities related to the environment and accidents. Although these latter two were ignored, their external cost would have to be fully internalised by relevant users. With respect to road accidents, over and above the Road Accident Fund levy included in the retail price of fuel, road users self-internalise part of the cost of accidents through voluntary vehicle accident insurance. Additionally, a proportion of accident costs are internalised through traffic fines for moving offences, such as speeding, drunken driving and driving an unroadworthy vehicle. Moreover, the cost of accidents in South Africa is calculated by using the gross-output method. In the case of fatal accidents, the present value of victims’ future loss of output is estimated. The economic cost of this tends to be overestimated in South Africa (given the country’s high unemployment rate and large unskilled population). With the large number of pedestrian fatalities in South Africa (especially in rural areas) the opportunity cost of fatal accidents is therefore overestimated.

5.2. DIVERSION TO SECONDARY ROADS

The transport system is ultimately a general equilibrium problem. Tolling one road in a wider network has an impact on the traffic flows of other roads (Leiman, 2003:130). As several toll roads in South Africa possess unmonitored and untolled side roads, the phenomenon of “rat-running” is likely to occur, i.e. motorists attempt to avoid road prices by driving on unpriced roads (Emmerink, 1998:44). Diverted traffic is likely to be heterogeneous in terms of vehicle composition. Rational drivers will divert from a tolled road to untolled side roads until the cost of doing so equals the cost of the toll on the primary road (Leiman, 2003:130). Specifically, truck drivers may divert under the explicit instruction of owners/managers or where incentives are positioned so as to encourage it.⁷⁰

These diversions have efficiency implications. Firstly, the untolled side roads are not designed to carry the level of traffic generated by the tolling diversion. This is likely to lead to, or aggravate, congestion on these secondary routes. The extra congestive effect of trucks *vis-à-vis* passenger cars has already been outlined, indicating the diversion of trucks onto alternative roads slows traffic flows.

Secondly, the side roads are designed with lighter pavement structures and are thus more susceptible to damage due to heavy axle loadings (Leiman, 2003). The diversion of heavy vehicles will therefore increase the maintenance costs of these roads, which are without the dedicated revenues for repair that toll roads enjoy.

Leiman (2003) provides an interesting case study for the diversion of heavy trucks from tolled highways to less suitable alternatives, using the example of the N1 Huguenot tunnel/Du Toit’s Kloof Pass. The Huguenot tunnel “is a high-speed dual carriageway that...cuts off 19 kilometres of steep, two lane, twisting and often mist-bound road with light foundations not intended for heavy vehicles.” (Leiman, 2003:131). The environmental impact assessment for further tolling on the N1 indicated that over 20 percent of trucks still made use of the alternate Du Toit’s Kloof Pass, despite the shorter, safer and quicker route that the tunnel offered drivers (Leiman, 2003:131). Leiman suggests that the continued use of the pass by heavy vehicles will lead to deterioration of the road condition, eventually pushing a number of users back onto the toll road, with the secondary road already having incurred the damage, as road users will divert until the marginal benefit of toll savings equal the marginal cost of the longer and poorer condition road (Leiman, 2003:131).

⁷⁰ Leiman poses the scenario that “[w]here a driver can pocket the money and the company picks up the incremental costs of extra mileage, additional repairs, time etc., the incentive to use side roads is even greater” (Leiman, 2003:130).

Keeler and Small argue that large-scale diversions from tolled roads are unlikely to persist in the long run, given the higher quality of service offered by toll roads over side roads (Keeler and Small, 1977:23). However, it is precisely because of diversions, which increase congestion and impart greater road damage, that secondary roads offer a lower quality service.

5.3. INCENTIVE TO OVERLOAD

Introducing tolls also creates the incentive for freight users to increase the weight of payloads above legally permitted limits. An insufficient network of operational weighbridges on the country's arterial roads hampers efforts to monitor and enforce mass limits. The authorities have noted that:

Failure to enforce limits on GVM and other road safety standards contributes to the unacceptable under-pricing of road usage. Effective enforcement is a prerequisite for the financial self-sufficiency of the road network. (Department of Transport, 1999:59)

Despite the higher legal GVM introduced in the 1990s, overloading remains prevalent in South Africa. In 1996, 58 904 heavy vehicles were tested at weighbridges and 33 percent found to be overloaded (CSIR, 1997:1). More recently, Leiman (2003) has indicated that around 15-20 percent of heavy vehicles are overloaded, causing as much as 60 percent of all road damage. The operation of the fourth power law indicates that the road damaging effect is particularly acute for overloaded vehicles.

Externality-correcting road prices to freight truck users will result in increased costs for freight truck operation, which will incentivise a greater fraction of freight trucks to overload as well as encourage already-overloaded vehicles to overload further. This is because the operator's objective function is to minimise the cost per ton transported. More particularly, the incentive is to minimise the toll cost per ton subject to the condition that the benefits of overloading (toll cost saving per ton) exceeds or equals the additional cost of overloading (risk of incurring a traffic fine, increased wear and tear on the vehicle, *etc.*). Absent adequate monitoring and enforcement, the risk of incurring a traffic fine is low and thus the expected benefit of overloading higher.

Any road pricing scheme that aims to recover road damage from users must therefore be matched by credible monitoring and enforcement efforts that disincentivise overloading.

5.4. PRICING IN GENERAL EQUILIBRIUM

The realisation that partial equilibrium solutions do not coincide with the general equilibrium case provides policy-makers with two road pricing options. Firstly, "quasi first-best" prices

may be set where indirect welfare effects are ignored and tolls are simply set to marginal external costs. However, quasi first-best tolls produce welfare gains that are less than second-best pricing and possibly induce welfare losses in certain cases (Verhoef, 2008:23).

Secondly, second-best prices may be used where first-best prices are optimally adjusted to account for distortions and inefficiencies in the economy. Second-best tolls may be higher or lower than marginal external costs, sometimes by substantial amounts (see, for instance, Rouwendal and Verhoef, 2006:108). Pricing rules here are more complex, requiring far greater information than simple marginal social costs and therefore increasing the likelihood of government failure. Moreover, second-best pricing may yield social welfare gains that are less, and sometimes far less, than the gains posited from first-best pricing.

While it is efficient to set a second-best toll, practically it appears that it is not a straightforward exercise to do so. However, attempts can be made to adjust prices as accurately as possible and use certain policies to limit distortions. These are explained briefly below.

5.4.1. Scope for second-best congestion charges

Consider road pricing on the highway with no price for other routes. While pricing use of the highway has the positive effect of reducing congestion on that route, it also has the effect of increasing congestion on secondary roads which are less capable of handling the traffic. Because of this negative side-effect, the second-best optimal level of the toll is below the marginal external cost on the highway (Verhoef, 2008:23).

This can be generalised in terms of the economic theory. Where two roads are substitutes and only one is tolled, the efficient second-best toll is less than the optimal toll if both were toll roads. However, the second-best toll becomes higher than the optimal toll if the two roads are complements (McDonald *et al*, 1999). Because the toll road is now underpriced with other roads not priced at all, welfare gains are substantially smaller than first-best pricing for all roads.

5.4.2. Scope for road damage-reflective charges

The use of technology may assist in circumventing distortions created by heavy vehicles in general equilibrium.⁷¹ Development of technology in recent decades means that it could be relatively cost-effective to equip heavy vehicles with in-vehicle remote communication devices (such as GPS or some similar technology) to accurately track distance and the type of road

⁷¹ Indeed, a National Treasury official has admitted that fitting vehicles with GPS transponders would be the best way of charging users and avoiding diversion onto secondary roads (Moore, 2012). This was stated with reference to the GFIP system and not specifically for heavy trucks.

travelled on. Only a small fraction of SA's total vehicle population (at most, 3.6 percent according to eNaTIS' December 2012 figures) would have to be fitted with such tracking devices.

This would also allow road agencies to observe where the heaviest travelled roads are, and allow them to respond to road maintenance issues if necessary. An important consideration for road damage is that damaged roads that are not repaired timeously deteriorate rapidly, leading to sharp increases in the cost of rehabilitation. Ideally the road agency would ideally want to price damage and use collected revenues to enact quick and appropriate repairs. SANRAL has indicated that maintenance delayed by 3 years from the point of ideal maintenance will cost six times more than ideal maintenance and 15 times more if delayed by 5 years (SANRAL, 2003:29). In other words, the marginal cost of repairs rises rapidly with deterioration (Ross, 2005). The lack of data on such issues has been cited by the DoT's Road Freight Strategy as a constraint on understanding the freight transportation industry and developing adequate policy responses (Department of Transport, 2011b).

Certain non-price policies to mitigate the effect of heavy vehicles on road damage are unlikely to make a difference. If trucks are regulated to stay in the slow lane, this has the same implication for road damage as a technical relationship requires all lanes must be resurfaced upon reaching a pre-determined surface quality (Newbery, 1989).

A more comprehensive weighbridge system is also required, on all arterial roads but particularly secondary roads (where construction specifications are not designed to deal with large volumes of heavy vehicles). This weighbridge network would assist in limiting diversions (in the absence of GPS technology and for weighbridges on secondary roads) as well as overloading, but chiefly allow for authorities to stop and weigh vehicles for the purposes of charging them for road damage.

There is, in fact, precedent in South Africa for the recovery of road damage from a specific class of vehicles. Mass-excessive "abnormal vehicles" (for example, mobile cranes) are currently required to purchase exemption permits based on a per-kilometre charge (derived from the vehicle's LEF and the mass fee) to cover the cost of the damage such vehicles impart on the road. These exemption permits come in the form of a trip or area permit, where the user is allowed a pre-defined number of kilometres of road travel during the permit validity period (Department of Transport, 2009b). While these permits serve as a proxy for average road travel and therefore damage, new technology allows for this to be done far more accurately, on a far wider scale, i.e. all vehicles with laden mass above some lower threshold.

The proposal to implement full road damage charges should not necessarily be introduced instantly. To align with the investment horizons of road carriers, road authorities should provide an appropriate time frame for road freight hauliers to comply with full road pricing and adjust their fleet renewal and composition plans accordingly.

Where road freight transporters face the full price for road use, they will face the correct incentives in terms of vehicle choice, loading factor, *etc.* Current toll pricing sets no incentive for truck firms to optimise social costs. As tolls are charged simply by axle number (and not weight), the incentive is to fully load fewer vehicles. The evidence is that road pricing for heavy vehicles in Germany resulted in a decrease in empty load vehicles by 7 percent and a fall in the use of less fuel-efficient heavy vehicles by 58 percent (Australasian Railway Association, 2010:46).

Almost a decade and a half ago, the DoT (1999:59) stated that “currently the price paid by road freight users for the use of the road does not cover the full costs of providing and maintaining the infrastructure and for externalities such as pollution and collisions” (emphasis added). Though the authorities are obviously aware that a degree of mispricing exists, there have been no efforts to correct these. In fact, increasing the vehicle mass limit in the 1990s only served to aggravate the under-recovery. A mooted amendment by the DoT in 2009 to reduce the legally permitted axle load on heavy trucks was met with vehement opposition from road carriers (see, for instance, Parker, 2009). The eventual proposed amendments published in 2012 contained no mention of revised axle limits (see Department of Transport, 2012b). The authorities either underestimate the severity of the mispricing or there are other factors at play. Indeed, one such factor may be lobbying power. The freight truck industry is well organised in the form of the Road Freight Association (RFA), having provided critical opposition to the proposed amendments to axle limits (see RFA, 2009).

6. A TALE OF TWO CASE STUDIES

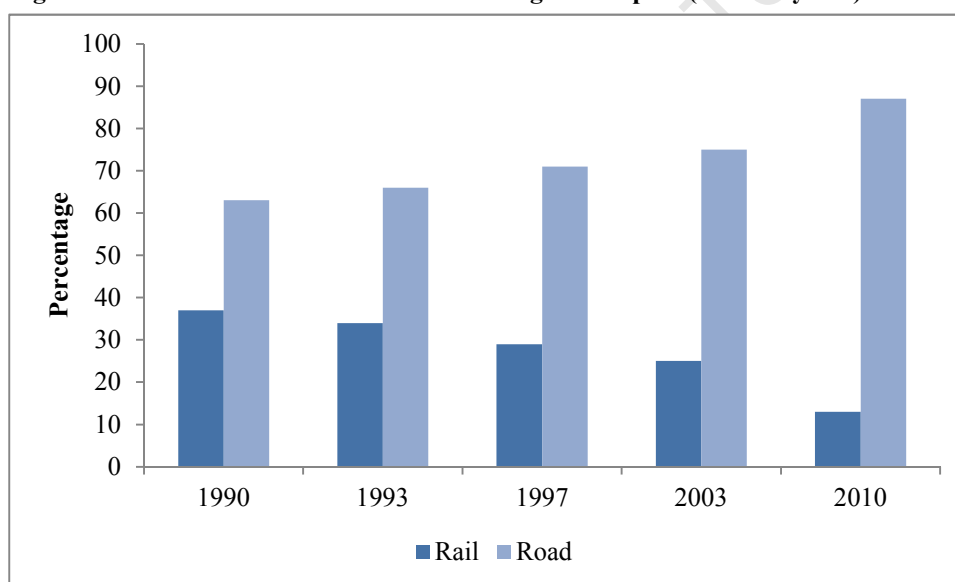
6.1. ROAD VERSUS RAIL – THE MARKET FOR LAND-BASED FREIGHT TRANSPORT

6.1.1. Freight transport in context

Consideration of the freight transport market indicates that the underpricing of road use implies subsidisation of road freight hauliers in South Africa. This has adverse effects for competition with the state-owned monopoly rail freight operator, Transnet Freight Rail (TFR).

Since the deregulation of freight transport in the 1980s, TFR has lost substantial market share in the freight haulage business to road carriers, with the majority of this loss along the country's main transport corridors.⁷²

Figure 9: Market shares for road and rail freight transport (selected years)



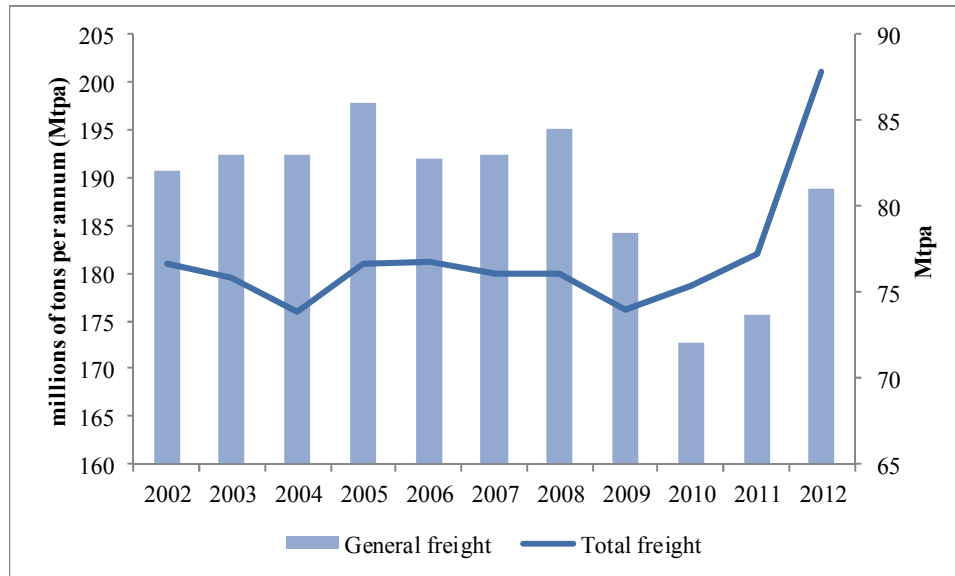
Source: CSIR State of Logistics reports, various years

In fact, the total tonnage carried by rail across Transnet's network was lower in 2011 (182 Mt) than 1990 (183 Mt). A similar situation can be observed for general freight. (See Figure 10.) Though absolute tonnage carried by rail has grown in recent years, this generally reflects increasing volumes of bulk mining commodities (iron ore, manganese ore and coal), the transportation economics of which strongly favour rail haulage.⁷³

⁷² Between these two time periods, freight was lost by rail along the rural and corridor parts of the network, while volumes grew in metropolitan areas and along bulk mining lines.

⁷³ That said, there is evidence that manganese is bagged and transported by road. See, for instance, Free State Department of Economic Development, Tourism and Environmental Affairs (2011) and Pallinghurst (n.d.). Additionally, it is well known that Eskom uses trucks for some coal transport. Indeed, observation of the condition of roads near the coal mines in Mpumalanga bear testimony to the road-damaging effect of laden trucks.

Figure 10: Historic rail freight volumes, 2002-2012



Source: Perkins (personal communication); Transnet annual reports, various years; DBSA (2012); Barrett (2005)
Note: Total freight volume on left-hand axis; general freight volume on right-hand axis.

6.1.2. Deregulating for competition

TFR's loss of market share is, in part, attributable to the long period of regulatory restrictions that limited competitive pressures and granted it an effective monopoly in the freight market.

The Regulation, Control and Management Act 22 of 1916 gave the South African Railways and Harbours (SAR&H) a mandate to provide rail infrastructure and services to industrial and agricultural users. Although SAR&H's mandate further required that it cover its costs from revenues, the government was allowed to fund any major projects that it deemed to be in the national interest (DBSA, 2012:12). The Motor Carrier Transportation Act 39 of 1930 followed, requiring that road haulage operators obtain licences from local road transport boards. In practice, however, these licences required the approval of SAR&H (DBSA, 2012:28). This gave SAR an effective monopoly over freight transport in South Africa. Some decades later, the Road Transportation Act 74 of 1977 further entrenched this monopoly, limiting road carriers to competing over distances of less than 80km where rail is ill-suited to compete.

Deregulation of the South African transport industry eventually took place in the 1980s. The promulgation of the Transport Deregulation Act 80 of 1988 provided road freight transporters with an opportunity to compete against the railways, which led to a sharp decrease in the capacity utilisation of the rail sector and a concomitant decline in investment in rail infrastructure (DBSA, 2012:12).

6.1.3. Addressing the operational efficiency of rail

Part of TFR's decline in market share is therefore a natural rebalancing of the competitive environment. This is especially true given the advantages provided by road freight are particularly pertinent in the modern economy as a result of just-in-time (JIT) production and the need for responsive distribution systems within urban areas.

What has subsequently driven the movement so strongly toward road freight is less clear. The putative poor management and operational inefficiency of TFR have been cited as primary factors behind this rapid and sustained shift. There appears to be a raft of evidence to suggest that historically the national rail operator (initially SAR, then Spoornet, and most recently TFR) has been plagued by poor operational procedures. This stems not only from its *de facto* rail monopoly and effective monopoly for all freight transport from the 1930s to 1980s but also the practice of using the entity to provide sheltered employment for "poor whites" during minority rule (Yudelman, 1983:237; Giliomee, 2003:340).

Undoubtedly, losses in freight volumes that have occurred in Transnet's General Freight Business (GFB) over the past two decades have been aggravated by poor customer service. Transnet itself has indicated that TFR is not offering the service levels that customers require, with the infrequent scheduling of trains for certain producers, repeated breakdowns of aging rolling stock, delays due to cable theft and infrastructure failure all hampering operations (Molefe, 2012). This has resulted in many firms making alternative arrangements with road transporters.

6.1.4. A fundamental mispricing in freight transport

While acknowledging these arguments, this paper posits that the shift in mode of freight transportation is, in no small part, due to the fundamental underpricing of road relative to rail. Although deregulation of road freight transport suggested an equal-opportunity market for road and rail, this mispricing means that this is not actually the case.⁷⁴

The rail operator not only covers the costs of running its haulage operations but is also responsible for track construction and maintenance (construction and damage costs). Any accidents that necessitate clean-up costs have to be covered by the operator (accident costs). TFR incurs and internalises the effect of any congestion on its network (congestion costs) (Leiman, 2003).⁷⁵

⁷⁴ The problematic nature of mispricing has plagued the rail sector for some time. See Van Walbeek and Pienaar (1996).

⁷⁵ The only notable road-type externality that TFR does not internalise is the environmental one.

Road freight carriers, on the other hand, share few such responsibilities and costs. They are not responsible for road construction and maintenance. The tolls that they pay capture neither the upfront cost of building roads to the specifications required by heavy vehicles nor the subsequent maintenance costs. In fact, the incentives of the roads agency and the road transporters are fundamentally misaligned. The absence of an incentive for individual truckers to help maintain the roads may lead to overloading. Such a situation does not exist for TFR.

In the face of such distortions, one cannot expect the market mechanism to allocate resources efficiently (Leiman, 2003:130). This goes some way to explaining the increasing share of freight haulage going to road transporters in South Africa. Firms, when making decisions upon transportation mode for freight, are facing pricing signals that steer them toward the road.

6.1.5. Estimates and effects of the mispricing

To get a sense of the current relative cost of freight transport, this paper compares estimates of the cost of shipping a container on the Durban-Johannesburg and Cape Town-Johannesburg routes for rail and road.

Interestingly, road freight transport is already more expensive than rail. (Prices should be seen as indicative only, as they are one-off cash quotations.) Building in the cost of unpriced road damage serves to sharpen the difference, with the price factor increasing from around two time to just under three for the CPT-JHB route and from 1.5 to just under 2 for the DBN-JHB corridor.

Table 18: Freight cost comparison

Mode	CPT-JHB (1408km)		CPT-JHB with road damage priced		DBN-JHB (569km)		DBN-JHB with road damage priced	
	Rail (13-20t)	R	8 083	R	8 083	R	7 003	R
Road (12-24t)	R	17 050	R	22 036	R	10 830	R	12 845
<i>Road:rail cost factor</i>		<i>2.1</i>		<i>2.7</i>		<i>1.5</i>		<i>1.8</i>

Source: Author's request for quotations from Transnet and Aplolor Express; Google Maps

Note 1: Freight origins are the harbours of Cape Town and Durban respectively. The destination is City Deep in both instances.

Note 2: Quotes for road and rail both contain delivery within 30km of City Deep. Road damage has not been included in either instance as the inclusion in both cancels out any relative effect.

Rail appears cheaper than road, and this cost difference is enlarged when pricing for road damage (without considering congestive and incremental construction cost externalities).

Two broad welfare effects arise from this mispricing. Firstly, there is a large social cost, borne by taxpayers and motorists. Given the size of the truck population, the absolute value of the cumulative externality is large. Motorists bear the cost of slow moving trucks on the road. For

damage and incremental construction costs, taxpayers are essentially subsidising heavy vehicles on non-toll roads, and light vehicles are providing the subsidies on toll roads.⁷⁶

Given the welfare costs, if rail were able to provide a perfectly substitutable service to road trucking then the current situation implies that taxpayers and motorists should be indifferent towards paying for trucks to sit idle with some subsidy transfer instead of being permitted to make use of the road.

Returning the road damage cost-recovery model from section 4.5, the table below provides indicative figures for the amount that representative heavy vehicles would have to pay in road damage tolls on an annual basis, given what they currently pay. These are therefore implicit subsidy values that the taxpayer and other motorists are providing to road carriers.

Table 19: Estimated annual road damage costs for representative heavy vehicles

Details	Absolute underpricing	
Minimum (Truck 8)	R	58 695
Maximum (Truck 4a)	R	578 996
Mean	R	246 438
Median	R	196 162

Source: Author's own calculations

Note: Full-load mileage of 112 500km assumed.

It may be argued that a subsidy for heavy trucks is warranted as their use contributes positively to the economy and that, given the operational inefficiency of rail, it may be road transport that enables the economy to survive and grow.

Though appealing, this argument may be invalidated by the impact the underpricing of road use has on the ability of the rail operator to adequately invest in capital expenditure. This is the second effect of the mispricing. The rail operator cannot generate sufficient revenue to invest in new capital and renew old capital equipment, and investment therefore declines to some lower, sub-optimal steady state. Indeed, there is evidence that there has been inadequate capital expenditure in TFR's operations in the last several decades, the fruits of which are seen in the complaints about poor operational efficiency. Transnet is currently attempting to rectify this situation by investing approximately 75 percent of its R300 billion capital expenditure programme, roughly R225 billion, to rail operations (Transnet, 2012:203).

This investment programme, however, poses an interesting complication. As part of Transnet's Market Demand Strategy (MDS), repayment of TFR's investment funding is premised on the outcome that it is able to attract freight from the road back to rail. However, Transnet's ability

⁷⁶ Congestion externalities arising from trucks, meanwhile, are borne by other motorists on both toll and non-toll roads.

to win back freight and thereby generate sufficient operating cash flows to sustain internal investment may be undermined by the distinct underpricing of road freight.

6.1.6. Effects of correct relative prices

There is evidence that Transnet's port and pipeline operations cross-subsidises TFR's current investment programme. To eliminate this distortion, it has been suggested that TFR's tariffs are raised to generate sufficient revenue or that the government provides direct subsidies to the rail operator (DBSA, 2012:15–16).

This obscures the fundamental problem. Re-jigging the pricing arrangement where road freight carriage was correctly priced would see relative prices increase to where Transnet's rail offering attracted greater demand. Additionally, it would provide room for TFR to raise its tariffs to levels that enable it to generate operational cash flows to finance capital expenditure programmes, thereby improving operational efficiency.

Coupled with an externality-corrective price regime, a non-price policy proposal that may facilitate the shift of freight from rail to road is the establishment of a more comprehensive system of intermodal hubs outside metropolitan areas. The advent of extensive containerisation supports the use of rail, where the long inter-city routes would be by freight train, and containers would then be moved by truck only from intermodal hub to destination. This would reduce the cumulative damage that trucks do, as there would be fewer on the road travelling shorter distances.

Government documents, such as the National Transport Master Plan (NATMAP), envisage a shift in passenger and freight traffic from road to rail (DBSA, 2012:55). This will not occur unless the incentives that users face through the price system fundamentally change. The current situation is neither equitable nor efficient for taxpayers, other motorists or other transport modes that compete against trucks, and represents a gross misallocation of resources.

6.2. THE GAUTENG FREEWAY IMPROVEMENT PROJECT

6.2.1. Background to the GFIP

SANRAL recently completed the Gauteng Freeway Improvement Project (GFIP), the largest project in the agency's history. While tolling has become increasingly prevalent on South African national roads in recent years, the introduction of the Gauteng Freeway Improvement Project (GFIP) marks a new direction for road pricing policy in South Africa.

The Gauteng government originally investigated tolling as an option for upgrading and expanding the province's highway network in the mid-1990s. Even at that time, the economic impact of traffic congestion was cited as a concern. Together with SANRAL, the provincial government considered implementing a concession project. An unsolicited bid was received and considered as a public-private partnership (PPP). Although initially given Scheme Developer status, the proposal was eventually found to be unacceptable because of the tariffs to be charged to users (DoT GFIP Steering Committee, 2011a)

Approved by Cabinet in 2007, the GFIP is a SANRAL-driven project that is anticipated to ultimately upgrade and construct 560km of freeways (SANRAL, 2012a). The plan has seen its first phase of 185km of freeway completed. The core improvements and benefits to the road system as a result of phase 1 are the addition of lanes (road widening) and interchange upgrades; lighting of all previously unlit road sections; and the introduction of equipment such as variable message signs, CCTV cameras and electronic traffic detection. Two further phases exist for the primary road network in the Gauteng region. Phase 2 of the GFIP is set to incorporate an additional 223km of the Gauteng province's potential toll roads.⁷⁷ Under phase 3 of the programme, 158km of new highway construction is planned (Van Niekerk, 2011).

To finance the construction of the first phase, SANRAL took on approximately R21 billion in borrowings against expected income streams from tolls to be collected under the system. In the face of public opposition to the pricing scheme, a special appropriation of R5.75 billion in the 2011/2012 budget was provided to ease the tariff burden on motorists (National Treasury, 2012). However, legal wrangling has delayed attempts by SANRAL to begin tolling along the system. An original date, April 2012, was set for tolling to begin. However, this was delayed in a case taken to the North Gauteng High Court by the Opposition to Urban Tolling Alliance (OUTA), where the project was put on hold subject to a judicial review. An appeal by SANRAL, National Treasury and the Department of Transport sought relief from the Constitutional Court, which granted the parties the right to continue with the tolling project with the review taking place concurrently. Regulations related to the practicalities of the GFIP toll scheme were gazetted and public consultations held thereafter. On 13 December 2012, the North Gauteng High Court dismissed the OUTA application for tolling on the GFIP to be set aside (Smith, 2012b).

As it currently stands, tolling along the GFIP route network may legally take place from early 2013, although OUTA is appealing the judgment of 13 December 2012 (Magubane, 2013).

⁷⁷ These include the N14 Krugersdorp Highway, a section of the M1 (between Woodmead to Sandton), the N14 Ben Schoeman Highway into Pretoria, N3 to Heidelberg, the R59 and N12 (Nancefield towards Potchefstroom) and the remaining untolled section of the N4 Pretoria towards Diamond Hill (Automobile Association, 2011).

6.2.2. Rationale for the GFIP

The rationale for tolling on the GFIP is two-fold, broadly falling under the DoT's user-pays principle:

Recovery of money spent on upgrading roads. As discussed in section 2, this is a standard motivation for tolling on the part of SANRAL and the government.

Congestion alleviation. The government seems acutely aware of the congestion problem on the Gauteng freeway system, and the National Treasury has advocated tolls on the GFIP as a means to alleviate this congestion. (See Gordhan [2012:25] and National Treasury [2012:121].) Comments made by officials further suggest that the National Treasury is of the view that tolls are the primary means of relieving congestion, and thereby improving economic efficiency through user prices (Moore, 2012). This is a notion that has not been raised in any previous motivations for tolling in South Africa.

6.2.3. Pricing of the GFIP

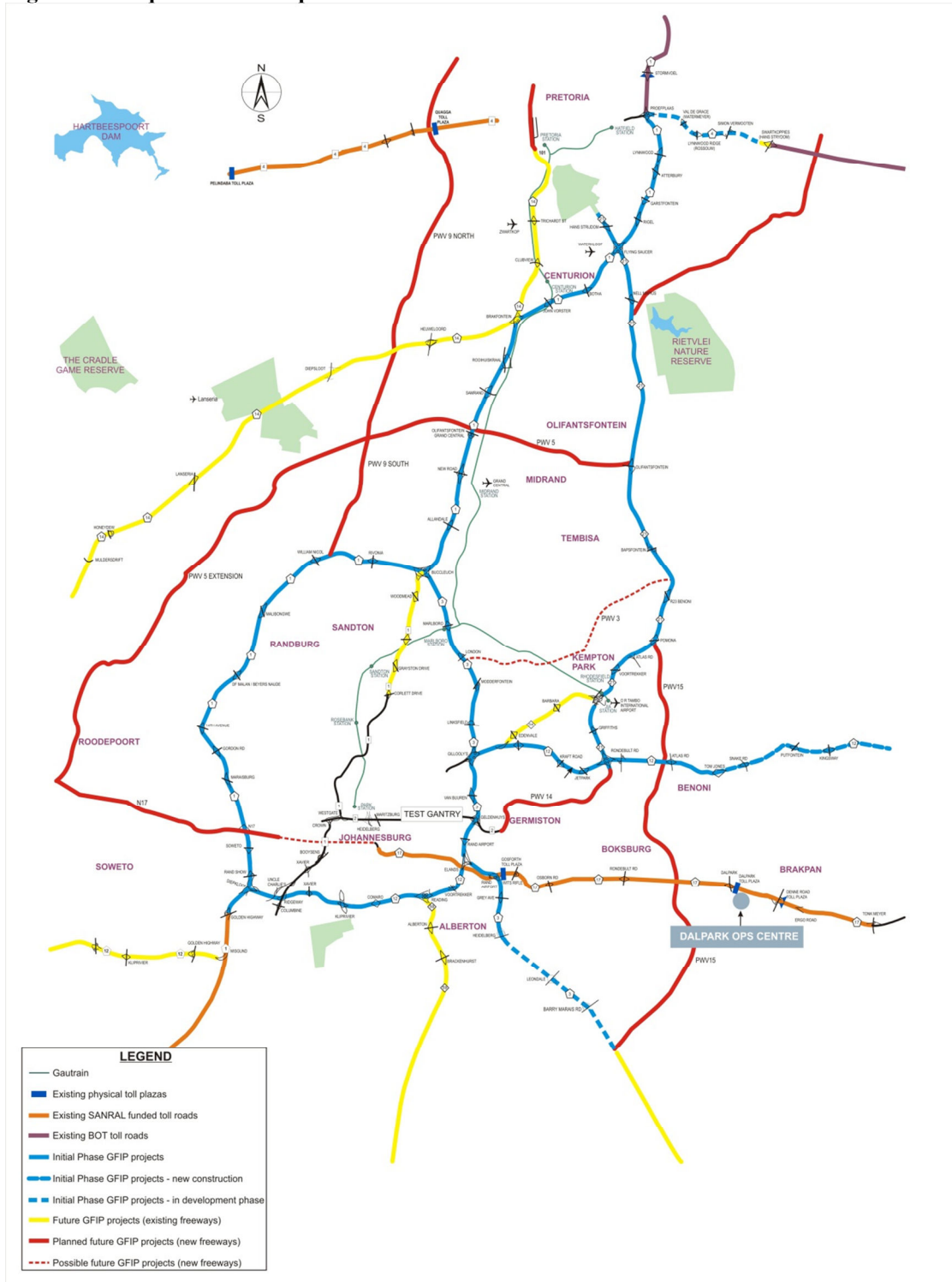
The GFIP provides for toll collection by electronic means only, under what is open road tolling (ORT) or "e-tolling". ORT, at least in the GFIP context, consists of large steel gantries that straddle all lanes in one direction of the road. Users of the system are able to acquire "e-tags", electronic devices that assist the authorities in verifying the user and collecting revenues. Frequent users are incentivised to obtain e-tags through considerable discounts on toll tariffs for holders of the tags. Gantries are equipped with CCTV cameras and electronic equipment that registers the number plate of a vehicle as it passes beneath the gantry. These gantries are strategically positioned at different points along the GFIP road network to ensure motorists pay for the extent of their road use.

Toll fees are charged on a per-kilometre basis. However, the structure of the system and positioning of the toll gantries is such that users are charged for "lumps" of their journeys upon passing beneath a gantry and, in certain instances, may not be charged for at all.⁷⁸

This paper will generally refer to e-tag toll rates as it is assumed that most users, over time, will respond to the discounts and obtain the device. Although there are calls from certain areas of civil society to boycott e-tag registration and ostensibly thereby make the administration of the system impossibly costly, the likely outcome will be that the private incentive to comply will overrun attempts for collective non-registration.

⁷⁸ This author notes that an irate motorist at the GFIP public information session of 15 November 2012 in Sunninghill indicated that a busy section of the N1 between New Road and Allandale Road was, in fact, completely free for use. (She was irate because the section of road she travelled on did have a gantry.)

Figure 11: Proposed route map for the GFIP



Source: DoT (2011a), GFIP Steering Committee Report

Tolls on the GFIP are differentiated for users on the following basis:

Vehicle class differentiation

Tolls charged under the GFIP follow similar SANRAL vehicle class distinctions to other toll roads, with vehicles placed into various classes. Notable, however, is that the classification system is entirely volumetric, or what the regulations call the “Volumetric Vehicle Classification System”.⁷⁹ Axle numbers therefore do not affect toll rates as is standard on other SANRAL toll roads.

Table 20: GFIP tolls per kilometre for e-tag holders

Vehicle type	Vehicle class	2011 tariff (original)	2011 tariff (revised)	2012 tariff (current)	Size of total saving
Motorcycles	A1	R 0.30	R 0.24	R 0.18	-40.0%
Light	A2	R 0.50	R 0.40	R 0.30	-39.4%
Medium heavy	B	R 1.49	R 1.00	R 0.75	-49.7%
Large heavy	C	R 2.97	R 2.00	R 1.50	-49.5%

Source: Government Gazette No. 35756 (25 October 2012)

Exemption is provided for public transport and emergency service vehicles in terms of Government Gazette No. 35755 (25 October 2012).⁸⁰

Toll rates are to be adjusted by the change in the consumer price index (CPI) on 1 March of every year, or by an amount the Minister deems fit.⁸¹

Time-of-day and day-of-week discounts

There are also discounts given to users of GFIP toll roads for certain times of the day and certain days of the week. Fundamentally, these discounts are to incentivise motorists to alter their travel patterns.

Table 21: GFIP time-of-day and day-of-week discounts

Time of day	Weekday		Weekend		Public holiday	
	A1 and A2	B and C	A1 and A2	B and C	A1 and A2	B and C
00h00-05h00	25%	30%	25%	30%	25%	30%
05h01-06h00	10%	25%				
06h01-10h00	0%	0%				
10h01-14h00	5%	20%	15%	25%		
14h01-18h00	0%	0%	25%	30%		
18h01-23h00	10%	25%				
23h00-00h00	25%	30%				

Source: Page 16 of Government Gazette No. 35756 (25 October 2012)

⁷⁹ Paragraph 4.1-4.3 of Government Gazette No. 35756, p. 14.

⁸⁰ This may have been seen by the authorities as a politically convenient decision given the strength of taxi operators as a (coordinated and un-coordinated) lobby group in South Africa. Interestingly, a similar decision in Hong Kong backfired when electronic tolling was brought to a stop because of driver protests against taxis being exempt (Maddison *et al*, 1996).

⁸¹ Section 5.22 of Government Gazette No. 35756

These discounts apply to all paying users (e-tag or not) and the discount is calculated as a percentage of the tariff that applies to a motorist. As the discount amount is the same for e-tag as it is for unregistered users, the real discounts for e-tag users are somewhat higher than what is displayed above.

Frequent use cap

E-tag holders may also benefit from frequency-of-use discounts. Each registered vehicle has a monthly toll amount that cannot be exceeded, regardless of distance of travelled.

Table 22: E-tag frequent user discounts

Vehicle class	Band	Marginal discount
A1 and A2	R400-R550	15%
	> R550	100%
B	> R1750	100%
C	> R3500	100%

Source: Page 17 of Government Gazette No. 35756 (25 October 2012)

6.2.4. Critical analysis of the GFIP pricing

Pricing congestion

Congestion tolling is a necessary price-based response to the major congestion found on Johannesburg's arterial highway system and interchanges, especially at rush hour during the week. The time-of-day and day-of-week discounts are presumably meant to capture the essence of the peak/off-peak congestion phenomenon and price it accordingly. This move by the authorities is commendable. However, it is necessary to critically analyse the system to determine whether the toll prices chose are socially optimal.

The welfare impact on the poor has been partly addressed by providing exemptions for public transport modes (buses and taxis). However, the limited public transport options provided in Gauteng means that the tolls may well have a negative welfare impact on those who would otherwise substitute away from private vehicle use.

Demand inelasticity is just one potential observation. Indeed, the GFIP tolls may just be a case of shifting congestion to secondary roads as motorists divert. The secondary road network of greater Johannesburg and its surrounds is such that numerous diversions from the tolled freeway are possible. Many of these secondary roads already experience heavy congestion, at peak hour and during the day, and traffic diversion will only serve to exacerbate this problem.

Where heavy vehicles divert they also impart significant damage to roads that were designed for lighter vehicles. While buses are exempt from toll fees and are therefore unlikely to divert, trucks are likely to divert. The Gauteng government and local municipalities therefore face the task of funding more extensive road repairs and maintenance. This is especially pertinent in light of the difficulties municipalities in Gauteng have had in attending to pothole problems on roads in recent years. The roads agencies in each of these spheres of government must fund the road repairs on secondary roads while toll revenue accrues entirely to SANRAL. Regardless of the long-run likelihood of diversions, the welfare impact of short-run diversions under the GFIP is likely to be relatively large.

Road damage and incremental construction costs

The GFIP Steering Committee noted in its final report that:

One of the key issues is how to pay for the rehabilitation and upgrading of roads. The most cost-effective way to pay for this rehabilitation and upgrading is through a combination of fuel tax and special levies for heavy vehicles. The special levies are necessary because, while heavy vehicles do the most damage to roads, these damages are not fully recovered in the fuel tax. The major constraint on the effective implementation of such a scheme is the financial policy on the part of government that fiscal integrity means that there should be no earmarking of funds. Hence all revenues raised, including the fuel tax, go into a common revenue fund and expenditures are made from this fund. (DoT GFIP Steering Committee Report, 2011a:14) (emphasis added)

While it is clear that the Steering Committee understand the road damaging effect of heavy vehicles, their conclusion misses the point of pricing road damage. The fungibility of contributions to the fiscus does not mean that the appropriate users should not be charged the correct price. Instead, it merely means that the contribution cannot be earmarked for road repair. This certainly does not make pricing the externality a worthless exercise.

The GFIP pricing policy for heavy vehicles further exhibits the misguidedness of SANRAL's notions for road-damage cost recovery. Instead of implementing a per-kilometre charge that prices the various externalities of heavy vehicles, SANRAL and the DoT have adopted a policy that, at times, appears ignorant of all but the congestive externality.

Firstly, the gazetted regulations stipulate charges by vehicle dimensions instead of by axle load (or even by a proxy for axle load, like number of axles). While this may be an attempt to provide a proxy for the congestive effect of vehicles, in any event, it outright ignores the impact of road damage.

Secondly, the monthly toll fee is capped at R3 500 for the “heaviest” (i.e. longest) vehicles. Trucks operators may simply see the charge as a lump-sum tax, as they may easily reach the maximum distance it implies (2 333km in a month). As such it will not positively affect their behaviour at the margin, for either the number of trips taken or the carried payload per trip.

Thirdly, although the peak/off-peak toll differential between heavy and light vehicles may adequately address congestive effects, it ignores the far more severe cross subsidisation between light and heavy vehicles for road damage and incremental construction costs. Further, large and medium “heavy” vehicles have disproportionately been the beneficiaries of tariff cuts, receiving larger toll savings than passenger vehicles.

Breakdown of toll revenues collected

The National Treasury indicates that the money raised from tolls on the Gauteng e-tolls will be ring-fenced and used for expenditure only related to the GFIP (Moore, 2012). Although the road life-cycle design is given as 24 years, toll pricing will remain indefinitely on the road (Van Niekerk, 2012).

Table 23 below details the constituent components funded by a representative Rand of GFIP toll revenue, based on the 2012 tariff structure.

Table 23: Breakdown of GFIP toll collections

Item	Percentage of total
Capital cost	28.9%
Interest	28.0%
Routine maintenance	14.9%
Violation costs	8.7%
Toll costs	17.1%
Freeway lighting	2.4%
<i>Total</i>	<i>100.0%</i>

Source: Alex van Niekerk and Marissa Moore (2012) GFIP public information session

Certain of these categories may be grouped and summarised as detailed below.

Table 24: Category breakdown of GFIP toll collections

Category	Percentage of total
Road benefits	46.2%
Interest	28.0%
Collection costs	25.8%
<i>Total</i>	<i>100.0%</i>

Source: Author's own calculation

Two points are notable from an examination of these tables. Firstly, an interest rate of 28 percent for the project is disconcertingly high, especially when one considers SANRAL's borrowing costs are marginally above the government's borrowing rate. Without access to the project's financial information, however, the reasons for this high interest rate are difficult to assess. The chief financial officer of SANRAL's explanation is that the magnitude is the result of delays in the implementation of tolling on the network (Mulder, 2012). One may have considered, however, that the capital transfer from National Treasury to SANRAL for the project would have partly alleviated the problems of compounded interest payments.

Secondly, collection costs are relatively substantial at 25.8 percent.⁸² This paper estimates that over the period required to pay back the capital amount of R21.7bn the cumulative collection cost is approximately R14.2 billion.⁸³ Assuming that the funding period exactly matches the design life of the road (24 years), this works out to be around R593 million per year for collecting annual revenues of roughly R2.3 billion. This amount is inclusive of the cost of collecting from e-tag and non-registered users. Presumably, the amount also includes a return for the toll management company, who, incidentally, will be paid approximately R1.7 billion per annum for managing the system (OUTA, 2013).

These collect costs amount to a significant amount of money. It has been suggested that part of the high costs is due to the selection of a foreign company (Krapsch) to provide the technology for toll collection. However, evidence from the tender process suggests that the winning bidder submitted a bid R2 billion lower than the next lowest offer (Sabinet, 2012).

Concluding remarks

It is evident that GFIP pricing is unlikely to be socially optimal, given the number of distortions that the tolling system creates as well as those externalities that the pricing policy ignores. Further, there may also be concerns about the cost of certain components of the system. When tolling for cost recovery rather than efficiency, the distortionary welfare effects of the system become even more pronounced.

⁸² Practically administering the system should not be much more costly than it is for Telkom to administer its billing system (Vickrey [1963] makes this point when considering the relative cost of administration for a telecoms operator and roads utility.).

⁸³ This assumes that the National Treasury grant of R5.75 was off-set entirely against the outstanding capital amount.

7. PERTINENT DISCUSSION AND POLICY POINTS

7.1. CONCERNS AND CONSIDERATIONS

7.1.1. Funding of national roads

Those opposed to toll roads have insisted that charging users for road access amounts to paying the state for an existing resource that has, ostensibly, already been funded.⁸⁴ This notion is based on the putative belief that the fuel levy and vehicle licence fees cover the costs of providing road infrastructure. As a result, the introduction of a road toll on an existing road is viewed as a tax increase or “double payment”.

While some level of collection costs must be seen as inevitable, relatively high toll collection expenses have prompted the suggestion that the fuel levy be used to raise the requisite money to pay for the infrastructure.⁸⁵ Given that the fuel levy is an existing funding mechanism, it is seen as a more appropriate and efficient method than ORT for funding the GFIP in particular and road infrastructure in general.

Its popularity is not only due to its simplicity and cost-efficiency (tax avoidance is difficult and administrative costs low), but also because of its marginal cost characteristics. To wit, users pay it for every kilometre travelled; the charge varies (to a limited extent) with the nature of the vehicle; it varies with the speeds at which vehicles travel and the manner in which they are driven; and it is easy to differentiate the charge between petrol and diesel-driven vehicles, which allows for the proxying, to a degree, of passenger vehicles and heavy vehicles respectively (Pienaar, 2005:12).

Because of the comparatively direct relationship between fuel tax and road usage, it provides an attractive basis for a road-use charging system. Moreover, the charge appears to garner public acceptability “because it is paid in small quantities at frequent intervals to satisfy immediate needs” (Pienaar, 2005:112).

However, it too has its shortcomings as an appropriate funding and pricing tool. Firstly, current fuel levy collections do not provide adequate revenue streams to cover total road expenditure in South Africa (DBSA, 2012:56). The levy already includes components for rebates to primary sector diesel users and sharing of the amount between the fiscus and metropolitan municipalities.⁸⁶ Even when excluding these two items, the absolute value of the fuel levy in

⁸⁴ An alternative way of positioning this is to say that people are opposed to having to pay for a good that was previously available at no explicit cost.

⁸⁵ This is both the general line of OUTA’s argument (see <http://www.ota.co.za/site/about-ota/why-we-oppose-e-tolling/>) and evidently the view of many motorists, as recorded at the public participation session in Sunninghill on 15 November 2012.

⁸⁶ For the 2010-2013 period, the fuel levy also contained a contribution toward the construction of the new fuel pipeline between Durban and Johannesburg (Department of Transport, 2011a).

2010 (the most recent data available) fell short of total road expenditure (Department of Transport, 2011a).⁸⁷

Secondly, an increase in the fuel levy to accommodate the road expenditure funding gap would induce distortions through cross-subsidisation. For a provincial-level charge, Gauteng motorists that do not use the highway system would be paying for something that only a fraction of users benefit from. This would be aggravated if the levy were increased on a national basis.

Thirdly, the fuel levy is not a mechanism that allows for accurate collection of externalities. For the road damage and incremental construction cost externalities, it is true that trucks use considerably more fuel than light vehicles. However, while fuel use and road use generally follow a linear relationship, the relationship between road use and road damage is an exponential one. It is also not clear that the fuel levy would be a useful means of pricing incremental construction costs. The discussion on congestive externalities also indicated that there was no relationship between fuel consumption and traffic flow. The fuel levy would therefore entail substantial cross-subsidisation between road users and would not improve current inefficiencies.

Fourthly, an obstacle that lies in the path of the fuel levy as a road funding mechanism is that it is National Treasury policy that all revenues that contribute to the fiscus are considered fungible, i.e. the fuel levy may not be ring-fenced for construction and maintenance purposes in democratic South Africa.

This, however, raises an interesting discussion point on road funding more generally. Though the budget process has to entertain competing demands, the expansion of the tax base in the past decade may indicate that there is, in fact, room for greater fiscal expenditure on road infrastructure. It is not necessarily the case that toll roads should be used to fund infrastructure. The existence of toll roads, and tolling as a funding option, may simply mean that toll revenues are actively displacing funding from the fiscus. Because roads have general economic benefits that push the PPF outwards, there is an argument to be made that fungibility is a less important principle than providing infrastructure finance from the fiscus. To achieve economic efficiency would require externality pricing when and where necessary, as discussed in section 5.4.

7.1.2. The desirability of private sector participation

Although SANRAL has increasingly made use of the private sector, it is worth assessing whether or not this is necessarily desirable when providing quasi-public goods. Should the

⁸⁷ This is for 2010, the most recent data this author could access.

government abdicate its position as infrastructure provider to private toll consortia and concessionaires, the resulting social welfare outcome may be no better than that of a governmental organisation and could, in fact, leave society worse off (Verhoef, 2008).

This is for several reasons, briefly discussed here. Firstly, there is the consideration of the comparative cost of funding. SANRAL is able to borrow in the open market for a relatively small spread above South African government bonds and then put the road construction out to tender with tolls set to cover interest and capital payments only. The government can currently borrow on long-dated bonds (30 years, which matches the standard concession period) at just over 8 percent.⁸⁸ The return on capital that a private BOT operator demands will undoubtedly be higher than this, purely given the higher cost of capital for private entities.⁸⁹ By utilising private sector funding rather than sovereign borrowings, the cost of capital for road construction and operation is likely significantly more than it could be under government.

Secondly, competition concerns arising from the allocation of property rights for the construction and operation of toll roads to private concessionaires are pertinent in a multitude of ways. Particularly, monopoly pricing power is relevant where the toll road is concessioned to a private operator.⁹⁰

Current SANRAL legislation allows for toll road operations on sections of the national road where no alternate route exists, a break with the founding legislation for the implementation of tolling.⁹¹ Hau states that:

the welfare gain from managerial efficiency due to private initiatives of road provision via increasingly popular build-operate-transfer (BOT) projects, for instance, should be measured against the welfare loss from monopoly abuse when parallel roads are non-existent. (Hau, 1998:68)

This gives rise to competition concerns, especially where a toll road with no alternatives is operated by a private consortium. The monopoly characteristics of roads in general, and those with no alternatives in particular, provide the operator with significant pricing power.

Though the implementation of BOT toll roads has been common since SANRAL's inception, it is important to note that functionally these are not so much BOT contracts as they are

⁸⁸ At the end of January 2013, the 30-year South African government bond yield was 8.21 percent (Bloomberg, 2013).

⁸⁹ Returns for construction engineering firms listed on the JSE Securities Exchange indicate that returns for the past ten years are in the range of 25 percent. As construction firms are free to invest their capital in those projects that yield the highest return, this is a fairly good proxy for the likely expected return.

⁹⁰ It may appear odd that this paper has argued that roads are underpriced but now states that monopoly pricing power may be a concern. This fear of "overpricing" pertains to light passenger vehicles, which are already overpriced along toll roads from externality analysis.

⁹¹ This was provided under the Second National Roads Amendment Act 79 of 1983.

maintain-operate-transfer (MOT). Indeed, Horak and Emery (2000) state that “[m]ost new toll road projects are in effect, not ‘greenfield’ projects, but upgrading and improvement of existing road infrastructure” (emphasis added). Toll operators are able to collect toll fees ostensibly under a BOT arrangement, though they do not have to incur the cost of constructing the road. Instead, tolls are charged and retained by the operator in exchange for maintenance as and when needed, until the road’s transfer back to SANRAL after a fixed period. This contractual arrangement may provide the private operator with room to negotiate toll prices that provide a higher-than-required return on capital invested.

Moreover, pricing power is a concern given the proliferation of open road tolling. Finkelstein (2009) finds that toll rates on electronic-collection roads are 20-40 percent higher than would be the case without electronic collection. It is also observed that the short-run elasticity of drivers decline when electronic toll collection is introduced (Finkelstein, 2009). In other words, drivers become less responsive to price increases when they are charged for road use by electronic means.⁹² These points are particularly pertinent given the imminent launch of the entirely electronically-collected GFIP and the likely roll-out of the devices along the country’s other toll roads.⁹³

Although toll rates must currently be vetted by SANRAL and the Minister of Transport, there are concerns with the transparency and equity of the process, in what public choice theorists would term “regulatory capture”.⁹⁴ The interaction between consortia members (predominantly engineering firms) and SANRAL may affect the ability of SANRAL to make objective and socially-optimal decisions.⁹⁵ The Minister of Transport and his advisors may not have the requisite technical knowledge (or specific information) to determine whether the toll prices applied for are adequate or excessive.

Finally, the propriety of SANRAL’s unsolicited bids process might also be considered. South Africa’s small, oligopolistic construction industry may raise competitive issues, especially when bids contain several construction companies. Unsolicited bids may also provide for perverse incentives. Private operators, in putting a bid together, may choose a stretch of road that does not contain any alternative routes. If this bid is successful, the private operator has obtained monopoly rights over that stretch of road. Upon receiving the bid, SANRAL has the incentive to develop the unsolicited bid as it reduces the network under management and therefore increases the available funds per remaining kilometre of road.

⁹² Finkelstein (2009) suggests this is because of what she terms as a lack of “tax salience” when these tolls are collected electronically.

⁹³ In fact, electronic collection technology has been used by SANRAL since 2003 (SANRAL, 2003:14).

⁹⁴ See, for instance, Stigler (1971).

⁹⁵ See allegations, for instance, from Rose, Hofstatter and Wa Afrika (2012).

Given that the government appears interested in pursuing private sector participation in road development, a challenge is therefore to design the process so that it transparently minimises the risks that the government (and society) face from rent-seeking private interest while not discouraging private sector participation. For example, game theoretical modelling may be useful to assist in the design of an optimal unsolicited bidding process. That is, “if the implementation of road pricing is connected with the privatisation of road infrastructure supply, there is good reason to develop smart concession mechanisms that aim to bring private capacity and toll choice closer to socially optimal levels” (Verhoef, 2008:15).

7.2. PROPOSING AN ECONOMIC REGULATOR FOR ROADS

The analysis and discussion above suggests that the application of road user charges may court controversy over a number of issues, including (i) the quantum of user charges, (ii) the basis upon which user benefits are measured (for example, distance travelled, number of passengers, load mass, *etc.*), (iii) differential treatment of different types of vehicles, (iv) a distinction between indirect taxes and road user charges, and (v) the proportional road cost responsibility of road users and non-road users (Pienaar, 2005:2). In light of these challenges, it is proposed that a road economic regulator (denoted “RER” from here) be established in South Africa.

This regulator would be an independent body responsible for the oversight of pricing on the arterial road network.⁹⁶ Though SANRAL has toll pricing as part of its current mandate, the incentives the institution faces, and the involvement of the private sector, means that the sector requires independent regulatory oversight. In the absence of appropriate regulation, welfare concerns exist.

An RER should therefore have the powers to oversee the road pricing regime. Roads network may be seen as a natural monopoly and appropriate usage pricing must be determined. Numerous authors support the view that the property rights for roads should vest with a public authority given the natural monopoly characteristics that roads possess (Newbery, 1989; Hau, 1998; Verhoef, 2008). As with many natural monopolies, appropriate monitoring of the sector requires an independent regulator.

In practice, the role of economic regulators for public utilities and natural monopolies is well established. In South Africa, two such regulators are the National Energy Regulator of South Africa, *inter alia*, responsible for approving electricity and commercial pipeline tariffs for the

⁹⁶ This would require that SANRAL and the Minister of Transport surrender their ability to vet toll prices.

state utilities Eskom and Transnet Pipelines respectively, and the Ports Regulator, which is responsible for approving Transnet National Ports Authority (TNPA) tariff applications.⁹⁷

There is also evidence that the concept of an economic regulator for roads has been established or proposed in several countries. In Australia, a relatively progressive road pricing policy exists where a newly established National Heavy Vehicle Regulator is in place.⁹⁸ In the United Kingdom, the Confederation of British Industry (CBI) recently publicly stated their support for the establishment of an independent road pricing regulator in that country (CBI, 2012).

Interestingly, the Australasian Railway Association has called for the establishment of a broader land-based transport regulator in Australia so as to address the distortions that exist between road and rail (Australasian Railway Association, 2010). There is no reason that this could not work equally effectively for South Africa. Such a regulator could also oversee TFR's operations, a sector that has received numerous calls for the establishment of an economic regulator given the monopoly incumbent. A land-based transport regulator would not only address the mispricing of road versus rail but could also improve the operational efficiency of freight rail operations in South Africa.⁹⁹

The RER would be an economically and socially effective institution for several reasons. Firstly, an RER would allow for prices to be set at optimal levels for economic efficiency rather than simply cost recovery. It would do so in a transparent manner and may increase public acceptance of road pricing as a result.

Secondly, it would enshrine the independence of the roads policy and construction process, shielding it from government interference and rent-seeking activities. In the first instance, this has been seen with the stop/start nature of tolling under the GFIP. Part of the problem is the scaling down of tariffs by the government in a potential bid to win over interest groups such as the RFA. In all likelihood, this industry association would oppose any externality-corrective pricing measures in a similar manner to which it successfully opposed a mooted reduction in axle limits for heavy vehicles and, more recently, railed against the cost of tolls for freight trucks along the GFIP (see Vermeulen, 2013). In the second instance, an independent regulator would act as a safeguard against private toll consortia who may seek to initiate toll roads at prices that are not socially optimal.

⁹⁷ NERSA's role in overseeing the tariffs for use of pipelines – an integral part of South Africa's transport infrastructure – illustrates that not only is there a South African precedent for independent regulation of utility infrastructure but a precedent for the regulation of transport infrastructure.

⁹⁸ See <<http://nhvr.gov.au>>. Admittedly, this regulator only pertains to heavy vehicles and not the road network as a whole.

⁹⁹ Rail network infrastructure (tracks, signalling equipment, *etc.*) is arguably a natural monopoly and should vest with the state, overseen by an appropriate regulator (an expanded RER). Track access could then be opened up for interested private operators, allowing them to compete against the incumbent TFR.

8. CONCLUSION

From a tumultuous past, where the institutional and legislative structure of national roads oscillated between centralised bureaucracy and autonomy, South African national road policy has galvanised under post-apartheid road agency SANRAL. Though toll roads were introduced on South Africa's national roads in the 1980s, they have grown significantly in kilometre length and funding size in the past fifteen years.

It is clear that a confluence of factors has contributed to the rise of toll roads in democratic South Africa. A strong desire by the government to develop a partially self-sustaining national road network has been a key factor. Tolls are thus a useful, and increasingly necessary, provider of funds for the construction and maintenance of the road network.

The government has also expressed a keen interest in the role of user charges to fund the construction and maintenance of the road network. Its user pays principle further aims at achieving economic efficiency by charging road users for the externalities generated as a result of road use, thereby internalising costs. Microeconomic theory dictates that it is efficient to price negative externalities that arise from the occurrence of an activity. Charges must be allocated to those who cause the damage and must differ by the level of externality generated.

The analysis in this paper shows that it is heavy vehicles that are responsible for an overwhelmingly amount of externality-generation. These vehicles do a greater quantum of damage to the road, make it necessary to build more expensive roads, and induce greater congestion by virtue of their size than other road users, disproportionately so during peak periods.

The magnitude of road freight trucks' external costs given the quasi-market prices these users face indicate that there is an absolute underpricing of road use for these vehicles in South Africa. On the grounds of the externalities considered in this paper, passenger cars should face a toll negligible in comparison to heavy vehicles. Instead, the factor for SANRAL's toll roads is in the range of 3 to 5 times greater for heavy vehicles than light vehicles – a magnitude that does not nearly equate the damage differential. Light vehicles should essentially only incur congestion charges. That they do not points to massive cross-subsidisation and that the relative price between light and heavy vehicles should be revisited.

Economic analysis of road use is premised on the basis that efficiency gains can be made by attributing relevant costs to each user. Efficiency is undermined where cross subsidisation takes place. In its policy document "Moving South Africa", the Department of Transport in 1999 clearly stated that "[c]ross subsidisation serves to distort price signals to customers and

cost signals to [transport] providers” (Department of Transport, 1999:52). The policy tool that the Department of Transport saw as most applicable to address this was the user-pays principle. However, in practice the aim of SANRAL has been to use tolls to fund road construction, not to enhance economic efficiency. As a result, road pricing in South Africa has been nowhere near achieving these stated goals and has, instead, been plagued by dramatic under-charging of heavy vehicles and ensuing cross-subsidisation.

With the network under management by SANRAL expected to grow substantially, and increase in importance with projects like the GFIP, further toll roads mean that it is critical for efficiency and equity that road use is correctly priced. Appropriate pricing will improve economic efficiency by reducing cross-subsidisation. It will also rationalise the choice of freight modalities in South Africa, with the likely effect that a greater volume of goods will be carried by rail.

Despite the welfare gains that the policy offers, one must be cognisant of the distortions that optimal road pricing may have. To assist in achieving second-best optimal pricing, policy proposals are given on how to better price road use and reduce unrecovered expenses on the road network. It is also proposed that an economic regulator for the road sector be established to develop appropriate pricing principles and provide adequate oversight for SANRAL and private sector participants, and shield investment and pricing decisions from lobby groups such as the RFA and bid consortiums.

Undeniably, further research is required before an advanced road pricing policy is fully rolled out for South Africa. This paper identifies that further refinement and quantification of the incremental construction cost externality is required, an externality that the literature appears to have overlooked when pricing road use. The dearth of road transport and freight data also hampers efforts, and should be urgently addressed by the relevant authorities. An increased commitment of resources to such research, and consequent policy action, will dramatically improve efficiency outcomes for the economy.

9. REFERENCES

Atkinson, A. and Stiglitz, J. 1980. *Lectures in Public Economics*. London: McGraw-Hill.

Australasian Railway Association. 2010. *Road Pricing Reforms in Australia: Why Road Pricing is Vital to Australia's Economic Prosperity*. [pdf] Available at: <<http://transportreform.org/file.php?fileID=26>> [Accessed 18 December 2012].

Automobile Association. 2011. *The AA Rejects the Recommendations of the Gauteng Open Road Tolling Steering Committee*. Available at: <<http://www.aa.co.za/about/press-room/press-releases/the-aa-rejects-the-recommendations.html>> [Accessed on 30 November 2012].

Ayres, I. and Levitt, S. 1998. Measuring positive externalities from unobservable victim precaution: an empirical analysis of LoJack. *The Quarterly Journal of Economics*, 113(1), pp.43-77.

Barnett, A. H. 1980. The Pigouvian tax rule under monopoly. *The American Economic Review*. 70(5), pp.1037-1041.

Barrett, J. 2005. *The restructuring of South Africa's state owned railway company, Spoornet*. [Presentation]. Available at: <<http://info.worldbank.org/etools/docs/library/211245/Spoornet%20presentation%20May11.pdf>> [Accessed on 03 February 2013].

Barro, R. J. 1990.. Government spending in a simple model of endogenous growth. *Journal of Political Economy*, 98 (5), pp. S102-S125.

Baumol, W. and Oates, W. 1988. *The theory of environmental policy*. 2nd ed. Cambridge: Cambridge University Press.

Behrens, R. and Wilkinson, P. 2001. South African urban passenger transport policy and planning practice, with specific reference to metropolitan Cape Town. *Urban Transport Research Group Working Paper 4*, Cape Town.

Bloomberg. 2013. *South Africa Govt Bonds 30 Year Note Generic Bid Yield*. Available at: <<http://www.bloomberg.com/quote/GSAB30YR:IND>> [Accessed 02 February 2013].

Buchanan, J. M. 1969. External diseconomies, corrective taxes, and market structures. *The American Economic Review*. 59(3), pp.174-177.

Button, K. 1998. Road pricing and the alternatives for controlling road traffic congestion. In K. Button and E.T. Verhoef. *Road Pricing, Traffic Congestion and the Environment*. Cheltenham: Edward Elgar, pp.113-135.

CBI. 2012. *Bold thinking: A model to fund our future roads*. [pdf] Available at: <http://www.cbi.org.uk/media/1775599/cbi_roads_funding_report.pdf> [Accessed 2 January 2013].

CCA Environmental. 2009. *Proposed N2 Wild Coast Toll Highway – Final Environmental Report*. Part E, Chapter 16. Available at: <http://www.nra.co.za/content/FinalEIR_Chap16_ConclsandRecs_Table16.1_16.16.pdf> [Accessed 4 January 2013].

COSATU. 2012. *COSATU to strike against e-tolls*. [press release] 17 April 2012, Available at: <http://www.cosatu.org.za/show.php?ID=6037>. [Accessed 20 May 2012].

CSIR. 1997. *The Damaging Effects of Overloaded Heavy Vehicles on Roads*. PAD27 – revised. Pretoria: CSIR.

CSIR. 2000. *Guidelines for Human Settlement Planning and Design (The Red Book)*. Pretoria: CSIR.

CSIR. 2004. *1st Annual State of Logistics Survey for South Africa 2003*. [pdf] Available at: <<http://www.csir.co.za/sol>> [Accessed 29 October 2012].

CSIR. 2012. *8th Annual State of Logistics Survey for South Africa 2011*. [pdf] Available at: <<http://www.csir.co.za/sol>> [Accessed 29 October 2012].

De Beer, M., Sallie, I., and van Rensburg, Y. 2009. *Revision of TRH 11 (1999 – 2000): Recovery of Road Damage – Discussion Document on a provisional basis for possible new estimation of mass fees – Under review for TRH 11 (2000) – Final summary report v1.0*. Available at: <http://researchspace.csir.co.za/dspace/bitstream/10204/2589/3/De%20Beer_2009_Update.pdf> [Accessed 15 July 2012].

De Kock, M. 1924. *Selected Subjects in the Economic History of South Africa*. Cape Town: Juta.

Department of Energy. 2012. *Diesel Price Margins*. [pdf] Available at: <http://www.energy.gov.za/files/esources/petroleum/July2012/Diesel_margins_guidelines.pdf> [Accessed: 16 July 2012].

- Department of Transport. 1996. *A future for roads in South Africa*. Pretoria: Department of Transport.
- Department of Transport. 1999. *Moving South Africa: The Action Agenda – A 20-year Strategic Framework for Transport in South Africa*. Pretoria: Department of Transport
- Department of Transport. 2005. *Strategic Plan 2005/2006*. Pretoria: Department of Transport.
- Department of Transport. 2006. *Road Infrastructure Strategic Framework for South Africa*. Pretoria: Department of Transport.
- Department of Transport. 2009a. *Strategic Plan 2009/10 – 2011/12*. Pretoria: Department of Transport.
- Department of Transport. 2009b. *TRH11: Dimensional and mass limitations and other requirements for abnormal load vehicles*. 8th ed. Pretoria: Department of Transport.
- Department of Transport. 2011a. *GFIP Steering Committee Report*. [doc] Available at: <<http://www.agbiz.co.za>> [Accessed 7 January 2013].
- Department of Transport. 2011b. Road Freight Strategy for South Africa. Final draft. Available at: <<http://www.agbiz.co.za/LinkClick.aspx?fileticket=eEDB8eDPB4Y%3D&tabid=113>> [Accessed 16 January 2013].
- Department of Transport. 2012a. *Abnormal Loads Exemption Permit Fees: Notice 297 of 2012*. Government Gazette No. 35234. Pretoria: Department of Transport.
- Department of Transport. 2012b. *National Road Traffic Act (93/1996): Amendment of National Road Traffic Regulations*. Government Gazette No. 35413. Pretoria: Department of Transport.
- Development Bank of Southern Africa. 2012. *The State of South Africa's Economic Infrastructure: Opportunities and Challenges 2012*. Midrand: Development Bank of South Africa.
- Downs, A. 1962. The law of peak-hour expressway congestion. *Traffic Quarterly*. 16(3), pp.393-409.
- Emmerink, R. 1998. *Information and Pricing in Road Transportation*. Springer.
- Finkelstein, A. 2009. *E-ZTax: Tax Salience and Tax Rates*. *Quarterly Journal of Economics*. 124(3), pp.969-1010.
- Floor, B.C. 1985. *The history of national roads in South Africa*. Stellenbosch: B.C. Floor.

- Free State Department of Economic Development, Tourism and Environmental Affairs. 2011. *Free State launches the containerised manganese project in Bloemfontein*. [press release]. Available at: < <http://www.edtea.fs.gov.za/wp-content/uploads/2012/05/MEDIA-STATEMENT-Transnet-DETEA-Signing-of-MOU-.pdf>> [Accessed on 27 January 2013].
- Freeman, P.N.W. 1981. *The recovery of costs from road users in South Africa*. DComm dissertation. Pretoria: University of South Africa.
- Gärling, T., Jakobsson, C., Loukopoulos, P., Fujii, S., 2008. Public acceptability of road pricing. In E.T. Verhoef, M.C.J. Bliemer, L. Steg, B. van Wee, eds. *Pricing in Road Transport: A Multi-disciplinary Perspective*. Cheltenham: Edward Elgar, pp.193-208.
- Giliomee, H. B. 2003. *The Afrikaners: Biography of a People*. C Hurst & Co.
- Gillen, D. 1997. Efficient use and provision of transportation with imperfect pricing: second best rules. In D. L. Greene, D. W. Jones and M. A. Delucchi, eds. 1997. *The full costs and benefits of transportation: contributions to theory, method and measurement*. Heidelberg: Springer, pp.193-218.
- Gillespie, T. D., Karamihas, S. M., Cebon, D., Sayers, M. W., Nasim, M. A., Hansen, W. and Ehsan, N. 1992. Effects of heavy vehicle characteristics on pavement response and performance. *Final Report – Project I-25*. University of Michigan Transportation Research Institute.
- Gordhan, P. 2012. *Budget Speech 2012*. Pretoria: National Treasury.
- Graaff, J. de V. 1968. *Theoretical Welfare Economics*. Cambridge: Cambridge University Press.
- Haiden, R.M., and Floor, B.C. 2003. Towards a policy on the imposition of road tolls in the City of Cape Town. *22nd Southern African Transport Conference*. Pretoria, South Africa 14-16 July 2003.
- Havenga, J.H. 2007. *The Development and Application of a Freight Transport Flow Model for South Africa*. PhD. Stellenbosch University.
- Hau, T. 1992. Economic fundamentals of road pricing: A diagrammatic analysis. *World Bank Policy Research Working Paper Series WPS 1070*. Washington D.C: World Bank.
- Hau, T. 1998. Congestion pricing and road investment. In K. Button and E.T. Verhoef, eds. *Road Pricing, Traffic Congestion and the Environment*. Cheltenham: Edward Elgar, pp.39-79.

- Hobart Houghton, D. 1967. *The South African Economy*. 2nd ed. Cape Town: Oxford University Press.
- Horak, E. and Emery, S. 2000. *Implementation of partnering in road construction*. [pdf] Available at: < http://reocities.com/athens/library/4702/papers/Partnering_WOAC_2000.PDF> [Accessed 4 January 2013].
- IEA. 2012. *Technology Roadmap: Fuel Economy of Road Vehicles*. [pdf] Available at: <http://www.iea.org/publications/fueleconomy_2012_final_web.pdf> [Accessed 8 January 2013].
- Jimenez E. 1995. Human and Physical Infrastructure: Investment and Pricing Policies in Developing Countries. In J. Behrman and T.N. Srinivasan, eds. *The Handbook of Development Economics*. Vol. 3B. Elsevier.
- Johansson-Stenman, O. and Sterner, T. 1998. What is the scope for environmental road pricing? In K. Button and E.T. Verhoef, eds. *Road Pricing, Traffic Congestion and the Environment*. Cheltenham: Edward Elgar, pp.150-170.
- Kalagadi Resources. 2012. *Manganese analysis: September 2012*. Available at: <<http://www.kalahariresources.co.za/index.php/media-room/54-manganese-analysis-september-2012>> [Accessed 22 November 2012].
- Keeler, T.E., and Small, K.A. 1977. Optimal Peak-Load Pricing, Investment, and Service Levels on Urban Expressways. *Journal of Political Economy*, 85(1), pp.1-25.
- Le Roux, H. 2005. State increasingly turns to tolling to deal with roads shortfall. *Engineering News* [online]. Available at < <http://www.corpcomm.co.za/index.php?pid=37&ct=1>> [Accessed 4 January 2013].
- Leiman, A. 2003. Efficiency and Road Privatisation: Bidding, Tolling and the 'user Pays' Principle. *South African Journal of Economics*, 72(2), pp.242-265.
- Lipsey, R. and Lancaster, K. 1956. The General Theory of Second Best. *The Review of Economic Studies*, 24(1), pp.11-32.
- Maddison, D., Pearce, D., Johansson, O., Calthorp, E., Litman, T. and Verhoef, E. 1996. *Blueprint 5: The True Cost of Transport*. London: Earthscan.
- Magubane, K. 2013. Outa still determined to stop implementation of e-tolling. *Business Day*. [online] 22 January. Available at: <<http://www.bdlive.co.za/national/2013/01/22/outa-still-determined-to-stop-implementation-of-e-tolling>> [Accessed 22 January 2013].

- Maluleka, K.J. 2008. *Transport Economic Regulatory Intervention in the Transport Infrastructure: A Public-Private Partnership Exploratory Study*. PhD. University of South Africa.
- McDonald, J.F., D'Ouille, E.L. and Liu, L.N. 1999. *Economics of urban highway congestion and pricing*. Boston: Kluwer.
- Mohring, H. and Harwitz, M. 1962. *Highway benefits; an analytical framework*. Transportation Centre: Northwestern University Press.
- Mohring, H. 1970. The Peak Load Problem with Increasing Returns and Pricing Constraints. *The American Economic Review*, 60(4), pp.693-705.
- Mohring, H. 1976. *Transportation Economics*. Cambridge, MA: Ballinger.
- Mohring, H. and Boyd, J. H. 1971. Analysing "Externalities": "Direct Interaction" vs. "Asset Utilization" Frameworks. *Economica*. 38(152), pp.347-361.
- Molefe, B. 2012. Transnet and Agri-logistics. *Agribusiness Indaba*. Johannesburg, South Africa 24 May 2012. Available at:
<http://www.abnproductions.com/events/agribusiness/Brian_Molefe_Agribusiness_Indaba.pdf> [Accessed 4 January 2013].
- Moore, M. 2012. Comments made on the Gauteng Freeway Improvement Project gazetted notices and regulations. [Public information session]. 15 November 2012. Sunninghill, Gauteng.
- Mulder, I. 2012. Comments made on the Gauteng Freeway Improvement Project gazetted notices and regulations. [Public information session]. 15 November 2012. Sunninghill, Gauteng.
- Munshi, R. 2012. Levy more expensive. *Financial Mail*. January 11-16 2013, pp.40-41.
- National Planning Commission. 2012. *National Development Plan*. Pretoria: National Planning Commission.
- National Treasury. 2012. *Budget Review 2012*. Pretoria: National Treasury.
- Newbery, D. 1988a. Road Damage Externalities and Road User Charges. *Econometrica*. 56(2), pp.295-316.
- Newbery, D. 1988b. Road user charges in Britain. *The Economic Journal*. 98(390), pp.161-176.

Newbery, D. 1989. Cost recovery from optimally designed roads. *Economica*. 56(222), pp.165-185.

OUTA. 2012. *Why we oppose e-tolling*. Available at: <<http://www.oua.co.za/site/about-oua/why-we-oppose-e-tolling/>> [Accessed 20 May 2012].

Pallinghurst. n.d. *Steel making materials*. Available at: <<http://www.pallinghurst.com/investments-tshipi-and-jupiter-mines>> [Accessed 27 January 2013].

Parker, A. 2009. DoT proposal sparks outrage. *Fleetwatch Magazine*. [Online]. Available at <<http://www.fleetwatch.co.za/magazines/Oct2009/18-DoT%20proposal%20sparks%20outrage.htm>> [Accessed 02 February 2013].

Paton, C. 2012. Minister sees role for business in roads, dams. *Business Day* [Online]. Available at: <<http://www.businessday.co.za/Articles/Content.aspx?id=166110>> [Accessed 29 February 2012].

Peninsula-Expressway Consortium. 2004. *Proposed N21 (R300) Cape Town Ring Road Toll Project..* Available at: <http://www.peninsula-expressway.org.za/draft_docs/Executive_Summary.pdf> [Accessed on 30 November 2012].

Perkins, P., Fedderke, J. and Luiz, J. 2005. An analysis of economic infrastructure investment in South Africa. *South African Journal of Economics*. 73(2), pp.211-228.

Pienaar, W.J. 2005. Road cost allocation and recovery. *Stellenbosch University Department of Logistics Working Paper WP1/03*. September 2005.

Pienaar, W.J. 2008. Economic evaluation of the proposed road between Gobabis and Grootfontein, Namibia. *South African Journal of Economics*. 76(4), pp.667-684.

Pienaar, W.J. and Nel, J.H. 2009. Mathematical principles of road congestion pricing. *ORiON: Journal of the Operations Research of South Africa*. 25(1), pp.45-51.

Pigou, A.C. 1920. *The Economics of Welfare*. London: Macmillan.

RFA. 2009. Reduced axle masses: RFA rallies industry together against proposed legislation. *RFA Update*. [Online]. Available at: <<http://www.rfa.co.za/rfa/downloads/members/news/newsletter/nov/rfa-newsletter-Nov09.pdf>> [Accessed 09 May 2103].

- Rose, R., Hofstatter, S. and Wa Afrika, M. 2012. Taking their toll: How a cosy club dominates SA's toll road empire. *Sunday Times*, 25 November, p1.
- Ross, D. 2005. On getting our money's worth from infrastructure spending. *SABITA Digest*, 2005, pp.13-24.
- Rouwendal, J. and Verhoef, E.T. 2006. Basic economic principles of road pricing: From theory to applications. *Transport Policy*. 13, pp.106-114.
- Roux, M. P. and Nordengen, P. 2010. Study on axle loads; turning circles; and lane widths for freight vehicles in sub-Saharan Africa. [pdf] Available at:
<http://www.trademarksa.org/sites/default/files/publications/Study_on_axle_loads_draft_final_report.pdf> [Accessed 13 July 2012].
- Sabinet. 2012. *Transport minister outlines Gauteng tolls tender process*. Available at:
<<http://www.sabinetlaw.co.za/transport/articles/transport-minister-outlines-gauteng-tolls-tender-process>> [Accessed 31 January 2013].
- SANRAL. 2001. *Annual Report 2001*. Pretoria: SANRAL.
- SANRAL. 2002. *Annual Report 2002*. Pretoria: SANRAL.
- SANRAL. 2003. *Annual Report 2003*. Pretoria: SANRAL.
- SANRAL. 2008. *Environmental Impact Report (draft)*. [pdf] Available at:
<http://www.nra.co.za/content/FinalEIR_Chap3_Fig3.1_3.3Tables3.1_3.4.pdf> [Accessed on 30 November 2012].
- SANRAL. 2009. *Declaration of Intent 2009-2012*. Pretoria: SANRAL.
- SANRAL. 2012a. *Overview of the project*. Available at:
<http://www.nra.co.za/gfip/index.php?q=con,23,%20Overview_Of_The_Project> [Accessed 9 March 2012].
- SANRAL. 2012b. *Strategic Plan 2012/13 – 2016/17*. Pretoria: SANRAL.
- SANRAL. 2012c. *Annual Report 2012*. Pretoria: SANRAL.
- SANRAL. 2011. *Latest GFIP News. July 2011*. [Online]. Available at:
<<http://www.nra.co.za/gfip/upload/LatestGFIPNews.pdf>> [Accessed 9 March 2012].

- Schussler, M. 2011. Consumer impact of the Gauteng Freeway Improvement Project. [pdf]. Available at: < <http://www.rfa.co.za/rfa/downloads/members/news/toll/20110411003.pdf>> [Accessed 12 April 2012].
- Small, K. A. 1992. Using the revenues from congestion pricing. *Transportation*. 19(4), pp.359-381.
- Small, K. A., Winston, C. and Evans, C. A. 1989. *Road work: a new highway pricing and investment policy*. Brookings Institute Press.
- Smith, N. 2012a. Cape Town council goes to court for Winelands toll review. *Business Day* [online]. Available at: < <http://www.bdlive.co.za/articles/2012/03/29/cape-town-council-goes-to-court-for-winelands-toll-review>> [Accessed 4 January 2013].
- Smith, N. 2012b. Outa weighs options after e-tolls defeat. *Business Day* [online]. Available at: <<http://www.bdlive.co.za/business/transport/2012/12/21/outa-weighs-options-after-e-tolls-defeat>> [Accessed 21 December 2012].
- Smith, N. 2012c. Second thoughts over split of Transnet Freight Rail. *Business Day*. [Online]. Available at: <<http://www.bdlive.co.za/business/transport/2012/09/17/second-thoughts-over-split-of-transnet-freight-rail>> [Accessed 15 January 2013].
- Solomon, V.E. 1983. Transport. In F.L. Coleman, ed. *Economic History of South Africa*. Pretoria: Haum.
- Stigler, G. 1971. The theory of economic regulation. *The Bell Journal of Economics and Management Science*, 2(1), pp.3-21.
- Transnet. 2012. *Long-Term Planning Framework*. Johannesburg: Transnet.
- Van Lingen, A. 1960. *A Century of Transport (1860-1960): a record of achievement of the South African Rail, Road, Air and Harbour Services*. Johannesburg: Da Gama Publications.
- Van Niekerk, A. 2011. *Gauteng Freeway Improvement Project GFIP: Current and Future Phases*. [pdf] Available at: <http://www.csir.co.za/Built_environment/Infrastructure_engineering/22RPF/04.%20A%20van%20Niekerk.pdf> [Accessed 19 December 2012].
- Van Niekerk, A. 2012. Comments made on the Gauteng Freeway Improvement Project gazetted notices and regulations. [Public information session]. 15 November 2012. Sunninghill, Gauteng.

- Van Walbeek, C.P. and Pienaar, W.J. 1997. Forecasting the demand for rail transport in South Africa. *Journal for Studies in Economics and Econometrics*. 21(3), pp.25-45.
- Verhoef, E.T. 2008. In E.T. Verhoef, M.C.J. Bliemer, L. Steg, B. van Wee, eds. *Pricing in Road Transport: A Multi-disciplinary Perspective*. Cheltenham: Edward Elgar, pp.6-28.
- Vermeulen, A. 2013. Association outlines year ahead. *Engineering News*. [Online]. Available at: <<http://www.engineeringnews.co.za/article/the-road-freight-association-outlines-its-year-ahead-2013-03-08>> [Accessed 09 May 2013].
- Vickrey, W.S. 1963. Pricing in urban and suburban transport. *The American Economic Review*. 53(2), pp.452-465.
- Vickrey, W.S. 1969. Congestion theory and transport investment. *The American Economic Review* 59(2), pp.251-260.
- Winston, C. 1985. Conceptual developments in the economics of transportation: an interpretive survey. *Journal of Economic Literature*. 23(1), pp.57-94.
- Winston, C. 1991. Efficient transportation infrastructure policy. *The Journal of Economic Perspectives*. 5(1), pp.113-127.
- Wray, Q. 2012. Government slashes Gauteng toll fees. *Business Report* [Online]. Available at: <<http://www.iol.co.za/dailynews/news/government-slashes-gauteng-toll-fees-1.1240409#.UQDZBieE2Ps>> [Accessed 23 February 2012].
- Yudelman, D. 1983. *The Emergence of Modern South Africa – State, Capital and the Incorporation of Organized Labor on the South African Gold Fields, 1902-1939*. Contributions in Comparative Colonial Studies, Number 13. London: Greenwood Press.

10. APPENDICES

10.1. APPENDIX 1 – MAJOR TOLL PROJECTS FOR SANRAL

Unsolicited bid proposals initially received were the R300 Ring Toll Road, the N1/N2 Winelands, the N2 Wild Coast Toll Road between East London and Durban, the N2 Garden Route between George and Port Elizabeth (297km), the John Ross Highway between Empangeni and Richards Bay, and the Super Highways Scheme between Johannesburg and Pretoria (Haiden and Floor, 2003). Two of these – the John Ross Highway and Super Highways Scheme – were subsequently cancelled (Le Roux, 2005).

It is also unclear at what stage the N2 Garden Road toll road is at. The SANRAL website contains a record of the process, with the last relevant date being 2009.¹⁰⁰ The *Civil Engineering Contractor* published an article in February 2012 stating that the toll road required only the authorisation of the Department of Environmental Affairs to proceed.¹⁰¹

SANRAL does have three major projects which are all unsolicited bids that it plans to take forward in the next several years. These are explained below to provide a view on the agency's proposed policy going forward. Three toll road PPPs are currently being considered (SANRAL, 2009:18):

- i. N2 Wild Coast Road
- ii. N1/N2 Winelands Road
- iii. R300 Ring Road

10.1.1. N2 Wild Coast Toll Highway (unsolicited bid)

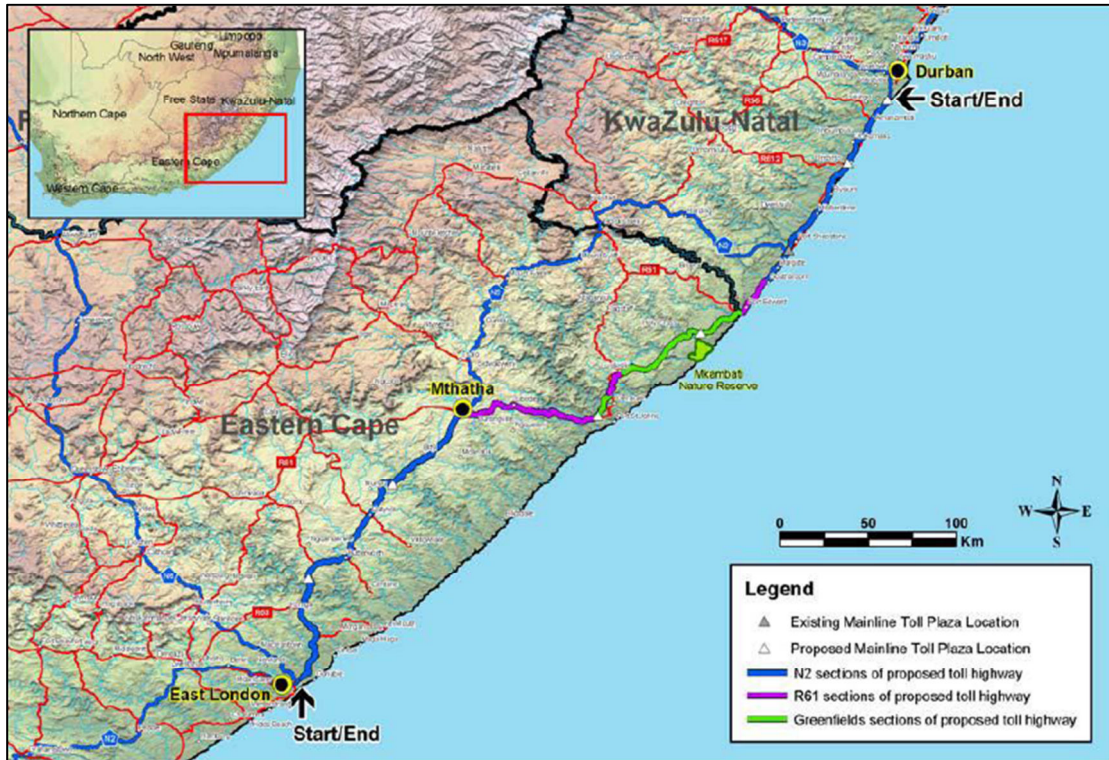
Situated under SIP3 (South Eastern Node and Corridor Development), the N2 Wild Coast project received environmental clearance from the Minister of Water and Environmental Affairs during 2011.

The proposed N2 Wild Coast Toll Highway includes seven mainline toll plazas covering approximately 559km (SANRAL, 2008:3-13). The planned construction is expected to reduce travel time between KwaZulu-Natal and East London and improve commuter and livestock safety (CCA Environmental, 2009).

¹⁰⁰ See < http://www.nra.co.za/live/content.php?Category_ID=167> [accessed 04 January 2013].

¹⁰¹ See < <http://www.myvirtualpaper.com/doc/brookepatrick/the-civil-engineering-contractor-february-2012/2012012012501/6.html#6>>.

Figure 12: Proposed N2 Wild Coast Toll Road



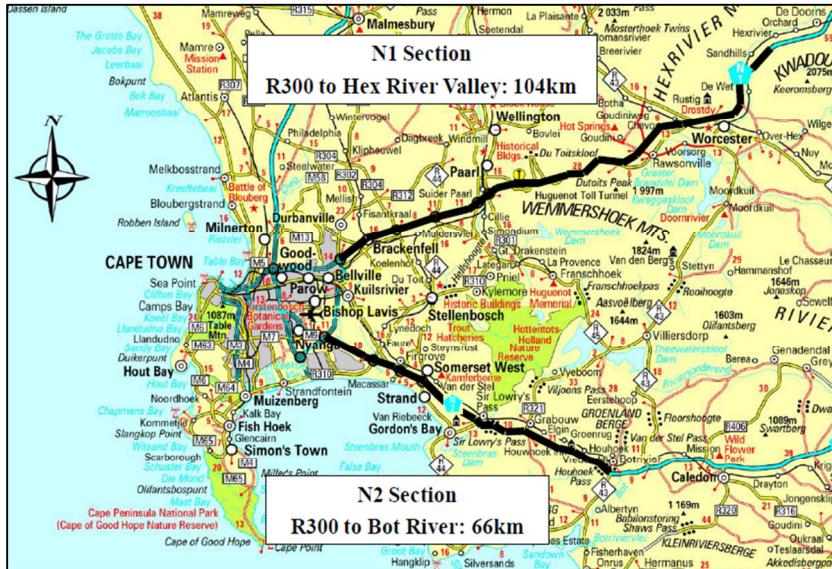
Source: SANRAL (2012)

The project was due to be tendered in 2012; however it is now on hold following government's reconsideration of the future of toll projects. In terms of the proposed concession contract, the concessionaire would be responsible for the operation and maintenance of the road for the duration of the concession contract, a thirty-year period (SANRAL, 2008:3-14).

10.1.2. N1/N2 Winelands Toll Highway (unsolicited bid)

The N1/N2 Winelands Toll Highway project is at the stage where the preferred bidder has been announced; however, contract negotiations between the proposed bidder and SANRAL have halted given government's current deliberation over its policy towards tolls as a means of financing infrastructure. Furthermore, the City of Cape has challenged the decision in the Western Cape High Court (Smith, 2012a).

Figure 13: Proposed N1/N2 Winelands Toll project



Source: SANRAL (2012)

10.1.3. R300 (N21) Cape Town Ring Toll (unsolicited bid)

The R300 (N21) Cape Town Ring Toll was originally conceptualised in 1995 to improve the road network in the greater Cape Town area. The primary idea is to provide a route to motorists that will bisect the highways and major arterial roads emanating from the city's central business district (CBD) (Peninsula-Expressway Consortium, 2004: 1).

Figure 14: Proposed R300 Ring Toll project



Source: SANRAL (2012)

It is envisioned that the project will cost R22 billion in total (at 2010 prices), including the construction of several new stretches of highway (SANRAL, 2012b:17).