




ARTICLE



<https://doi.org/10.1057/s41599-020-0508-2>

OPEN

A stochastic prediction of minibus taxi driver behaviour in South Africa

Jan Schlüter ^{1,2✉}, Manuel Frewer^{1,3}, Leif Sörensen^{1,4} & Justin Coetzee⁵

For many emerging and developing countries, urbanization and demographic changes require a remodelling of transportation systems and networks. Paratransit, with its demand responsive characteristics of being fully flexible, often represents the backbone of the transport sector while being an informal and, therefore, fairly unorganised system with little knowledge about working routines of its workers. As an example, this paper analyses the minibus taxi industry in South Africa based on data from Rustenburg, a mid-sized city, to gain valuable insights into the working behaviour of minibus taxi drivers by taking an evidence-based and data-based approach. Therefore, the principle of reference-dependent preferences is applied to the data to investigate the labour supply choice of taxi drivers, since sheer profit maximization has proven inadequate for the analysis here. Results indicate, that working hours for South African minibus taxi drivers are likely to be in the 10–12 h bracket with a revenue-dependent additional hour accounting for the revenue performance of that day. As a main finding however, South African minibus taxi drivers cannot clearly be attributed to either reference dependence or profit-maximization behaviour.

¹NGM, Department of Dynamics of Complex Fluids, Max Planck Institute for Dynamics and Self-Organization, Am Fassberg 17, 37077 Göttingen, Germany.

²Institute for the Dynamics of Complex Systems, Faculty of Physics, Georg-August-University of Göttingen, Friedrich-Hund-Platz 1, 37074 Göttingen, Germany.

³Chair of Statistics, Department of Economics, Georg-August-University of Göttingen, Humboldtallee 3, 37073 Göttingen, Germany. ⁴Thinktank of Aeronautics, Aerodynamics and Aerospace Technology, Marschblick 5, 25866 Mildstedt, Germany. ⁵GoMetro, Cape Town, South Africa. ✉email: jan.schlueter@ds.mpg.de

Introduction

Shared taxis and demand responsive transport (DRT) are believed to be one of the major means to cope with the mobility needs that follow from the rapid urbanization and its resulting traffic congestion caused by more and more individual private car usage. At the same time, the loss of infrastructure in rural areas that stems from population movements away from the periphery towards the urban cores leaves the residual population with few means of transport, as any public service would lack proper utilization rates. DRT could be a solution to these problems, as it promises to efficiently combine the trips with low passenger numbers into one single trip. The route is calculated (or chosen) in such a way that it combines the maximum number of passengers in one single trip, at the same time keeping the driven distance as low as possible (see Brake et al., 2004).

In many developing countries, such as South Africa, a mode of transport that has demand-responsive characteristics has been meeting the mobility needs of the lower and lowest income brackets for decades already. Substituting for governments that are either unable or unwilling to offer subsidised means of public transportation in the periphery of metropolitan areas, a privately owned mode has developed, delivering mobility to the less well-off. *Paratransit*, or in the specific case of this paper *minibus taxis* are often the only mode of transport available to the poorest parts of the population living in the periphery of metropolitan areas to access jobs and services located in the centres. Thus, paratransit, e.g. *Matatus* in Kenya and Uganda, *Jeepneys* in the Philippines or *Colectivos* in South America, represent a substantial infrastructure with employment opportunities for many people, as well as granting access to employment for many more. Consequently, the functioning of these paratransit services as well as the understanding of underlying current dynamics has a global impact for shaping this sector in the future. The minibus taxis operate without any technical support (e.g. smartphone apps) and information is only dispersed through word-of-mouth. Payments are exclusively made in cash. Yet, in many ways this paratransit emulates shared taxi schemes and is thus demand responsive in nature. Buses mostly (but not exclusively) start from widely known points in the city to start their journey, picking up passengers along the way and letting off passengers when and where they demand it. Most importantly yet is the fact that the South African minibus taxis are entirely privately organised, not receiving, with few exceptions, any government subsidy (see Neumann et al., 2015). This also holds for many other developing countries. At the same time, almost the entire sector operates within the informal economy and does not strictly adhere to regulations, labour, or safety standards and does not pay taxes (see Woolf and Joubert, 2013). In this regard, analyses of South Africa's minibus taxis can hold valuable lessons for the organization of shared taxi schemes elsewhere where formalization is also yet to be realised. In South Africa the paratransit sector has been an autonomous response of the Black and Coloured communities to the lack of development and planning of the transport network for Black and Coloured main residential areas by the government in apartheid times known as 'spatial segregation' in the literature (see for example Mossay and Picard, 2019). The minibus taxi industry as it is known in South Africa retains a market share of around 60% in the public transportation sector until the present day and is a vital part of the South African transportation ecosystem (see NHTS, 2013). However, the sector is commonly known to cause a wide array of problems ranging from congestion, due to the amount of buses, to outbursts of violence between different taxi operators on the streets (see Braumann and Haferburg, 2010). This has caused policy makers to attempt to regulate a sector that has traditionally been operated

to a maximum degree of decentralization, albeit with mixed success. The major problem with regulating the industry lies in its informal and by extension chaotic nature (see Neumann et al., 2015). Thus far any regulation has been based on anecdotal evidence or self-reporting, which could explain many of the unsuccessful policies. Until recently, no quantitative surveys on minibus taxi operations have been carried out in order to establish operating and business characteristics of the minibus taxi sector.

This paper seeks to establish a view on the minibus taxi sector by taking a more evidence-based and data-based approach as has been the case so far. This is made possible through the valuable contribution of the Cape Town-based company GoMetro, which generously has made their survey data available to the authors. The data set comprises data on minibus taxi trips, around 8000 in Rustenburg, South Africa (about 100 km west of Pretoria) and around 4000 in Cape Town. This paper takes a more specific look into the behavioural features of the drivers of the minibus taxis, as their style of conduct is often described as erratic and a main cause of concern that passengers have against minibus taxis (see NHTS, 2013). Secondly understanding the incentives and drivers' reactions to them could be considered a valuable asset in further organizing and formalizing the sector, for example through hypothetical ride-hailing applications that encompass the minibus taxi industry. As an example: Understanding when a driver is most likely to start and end his or her trip or working day will help to predict driver availability and would thus aid the smooth functioning of such an application. Concretely, this paper will quantify certain points (in time spent driving and revenue earned) where drivers decide to cease to drive or to continue. Based on the global similarities in paratransit services, these findings contribute not only to the understanding of South African but international paratransit services in developing countries.

This paper is organised as follows: section "State of research" gives a brief overview over the current state of research and literature that described the minibus taxi industry, as well as the reference-dependent preferences and its applications. Section "Theoretical framework" draws up a theoretical framework for the behaviour of minibus taxi drivers under reference-dependent preferences. Section "Results" presents and discusses the results and section "Conclusion" concludes this paper.

State of research

The research centred around the so-called *minibus taxi industry* located in South Africa mainly focuses on its politically controversial nature and the policy attempts to either formalise (e.g. Barrett, 2003; Schalekamp and Behrens, 2010; Woolf and Joubert, 2013; Joubert, 2013; Germinah and Chiloane, 2016) or replace and integrate it with more formal bus systems (e.g. Del Mistro and Behrens, 2015; Venter, 2013; McCaul and Ntulum, 2011), partly finding mixed effects of the implementation of high capacity bus routes on the demand for minibus taxis. Schalekamp and Behrens (2010) highlight in the context of any formalization in South Africa that relevant stakeholders have to engage in joint efforts to improve the transport system. The minibus taxi industry proved resistance against changes to their industry in the past. The highly ambiguous taxi recapitalization programme has widely been criticised for its top-down approach, its lack in funding and proper incentives while being too demanding on the sector operators with respect to the conditions (see Schalekamp and Behrens, 2010; Woolf and Joubert, 2013; Joubert, 2013; Germinah and Chiloane, 2016). In this context, Barrett (2003) emphasises on the importance to formally organise the employees

in the industry to have a joint standing against the government and the business owners. Generally, the South African governments approach towards the sector has been criticised as too heavy-handed and mainly focused on ‘cleaning up the streets’ to convey the image of a developed country (see Braumann and Haferburg, 2010; Venter, 2013; Barrett, 2003). Few published works have structurally investigated the supply side of the minibus taxi industry with the exception of Barrett (2003) and Woolf and Joubert (2013). To the knowledge of the authors, even fewer works have a quantitative founding, Del Mistro and Behrens (2015) is an exception as demand side data is used. Due to the informal status of the sector, quantitative data are naturally hard to come by. On the other hand, there is a quite a wide literature on the so-called *paratransit* sector. While usually being defined in developed countries as a transport method for the disabled, in developing countries, the term refers to all kinds of transport that are comparable to the minibus taxis already introduced. The main characteristics of this sector are non-availability of a schedule and a fixed route, that is not organised in a central way. These papers mainly deal with the sector from a city planning perspective, whether to combat or facilitate the sector and how the situation was handled in different countries and cities e.g. Bogota, Columbia; Mexico City, Mexico; Jakarta, Indonesia among others (see for example, Cervero and Golub, 2007; Fernández et al., 2007; Jokinen et al., 2001; Nguyen-Hoang and Yeung, 2010; Phun and Yai, 2016; Schalekamp and Behrens, 2009).

This paper’s theoretical foundation is based on the pivotal and Nobel laureated work of Tversky and Kahneman (1991, 1979). In brief, the psychological findings of Tversky and Kahneman have shown that the applicability of the neoclassical expected utility approach, meaning the assumption of risk-neutral agents that assess risk rationally, is only limited. They found that most study participants preferred small but relatively secure gains over high but improbable ones (while both exposed an equal expected value). This became known as the prospect theory. Later, the theory of reference-dependent preferences was developed based on this (see Thaler, 1980; Köszegi and Rabin, 2006). Here, an agent has a reference level that she or he will try to attain, and which is the main driver of her or his utility. Köszegi and Rabin (2006) address in their work the role of endogenous and exogenous reference levels, e.g. a driver is likely to work more (or longer hours) when the predicted revenue is higher (compared to endogenous reference levels) but may work less if higher wages were not predictable when deciding on whether to work or not. Empirically these findings have been tested on taxi drivers, which are a natural choice for studying the underlying dynamics of labour supply as they are largely self-employed and can decide on an individual basis on their work times. Moreover, their work schedule is tightly tracked. While earlier research (e.g. Camerer et al., 1997, Farber, 2005, 2008) depend on smaller samples of single taxi companies newer contributions take into account the full scale of the New York City Taxi and Limousine Commission (TLC), which, since it has made it mandatory for all taxis to track all their trips electronically, collected millions of observed trips. Recent research (e.g. Martin, 2017) is based on a dataset of 116 million trips. Mostly employing varying types of regression spline models (as does this work) a number of well received investigations have found a limited applicability of reference-dependent preferences. Some investigations attempted to measure whether the wage elasticity is positive (thus working more when there is more to get according to the neoclassical model) or negative (working less when wage is high, as the reference income is earned faster, e.g. Camerer et al., 1997, Chou, 2000, Farber, 2015). Others focus on interpreting driver’s utility via their stopping behaviour (e.g. Farber, 2005, 2008; Crawford and Meng, 2011;

Martin, 2017), or earning shocks, such as tips (e.g. Schmidt, 2018; Thakral and Tô, 2017).

Theoretical framework

Since the inception of risk-averse preferences by Tversky and Kahneman (1991) much research has been adjusted to the empirical reality that economic agents do not seem to follow neoclassical styled rationality. A widely known example is the so-called *endowment effect*, a phenomenon that has shown through experiments that agents value goods in their possession higher than equivalent goods that are available for purchase to them, an assessment that seems to be bar of any economic rationality (see Thaler, 1980). This has led to the development of reference-dependent preferences, which are explained in the following section. This section seeks to establish what is consensus in the research that has applied reference-dependent preferences to labour supply choice of taxi drivers and explores to what extent these results can be extended to the taxi schemes in South Africa and what potential adjustments would have to be made in its particular case for minibus taxi drivers.

Reference-dependent preferences. The principle of *reference-dependent preferences* is founded on the already mentioned endowment effect. A string of experiments with graduate students of economics have found that economic agents (even when faced with learning opportunities and real market environments) will value something that they possess much more than something they can purchase. In other terms: Their willingness to accept (WTA) a financial compensation for selling a good in their possession is much higher than their willingness to pay (WTP) for a strictly equivalent good that they are to purchase. Many experiments have thus found that the quoted median selling price in these experimental markets is twice as high as the quoted median buying prices. The conclusion from these experiments was that these symptoms must be due to risk-aversion (see Kahneman et al., 1991). Utility, it has been argued, is thus not necessarily derived from owning more of one good, but mainly due to the departure from a neutral reference level. Another pivot was that departing from this reference towards a more negative result is perceived much greater a loss, than an equivalent gain in endowment (and hence utility) compared to what the reference would have had. Based on these experimental foundations the concept of reference-dependent preferences has been drawn up, which is taking into account the observation that agents perceive a loss as much more grave than an equivalent gain, hence marginal utility (e.g. in income) is higher in an area of loss than marginal utility is in an area of gains. This means that a utility function will have two different ‘sets’ of slopes and hence a ‘kink’ located exactly at the reference level (see Tversky and Kahneman, 1991, 1979).

Applicability towards South Africa Dataset. A wealth of research has been done to empirically estimate whether taxi drivers follow a neoclassical model of labour supply, which requires that taxi drivers follow a form of inter-temporal substitution (e.g. driving more on a ‘good day’ and pass by leisure and vice versa for ‘bad days’), or whether they follow a model that resembles what was described as reference-dependent preferences further above and have a certain amount of income in mind that they want to reach on any given day. This has implications far beyond the realm of labour- and microeconomics. Neoclassical approaches imply strictly rational preferences of individuals under the assumption of perfect information. Thereupon, individuals maximise their utility. Hence under the neoclassical model, a driver will continue to supply labour as long as profits

can be realised as the utility for supplying labour is higher than for any alternative, e.g. leisure activities. Otherwise, when profits are no longer feasible ('bad days'), opportunity costs of labour supply increase compared to leisure activities which results in withdrawing labour from the market (inter-temporal substitution). Many of the workhorse models of macroeconomic theory depend on the assumption of inter-temporal substitution of labour supply that is challenged by reference-dependent preferences.

Minibus taxi drivers in South Africa are a natural choice for studying the underlying dynamics of labour supply. They work in a framework similar to self-employment as they rent the taxi from the owners and pay them around 30% of the daily revenues. Therefore, they are incentivised to maximise passenger turnover and hence the number of trips made. Furthermore, they decide autonomously on whether to continue working or ending their shift as they are (mostly) not part of a fixed work time scheme.

The empirical test for reference dependence can be divided into different approaches. For one, the measurement of wage elasticities via a logistic regression framework: Neoclassical theory would suggest a positive wage elasticity of supply. However, if reference dependence is relevant, a negative elasticity of around -1 should be observable, as an increase in wage would directly lead to less labour supply since reference income can be reached faster and with less time spent working (see Camerer et al., 1997; Martin, 2017). Another approach is to measure different sets of brackets of income (e.g. 150\$– 175\$) on a latent variable that measures the 'hazard of stopping', meaning to withdraw labour supply by ending the trips. This approach has the serious drawback that any income bracket is arbitrarily chosen (see Farber, 2005, 2015; Martin, 2017). A third method would be to directly assess the relevance of the difference between a reference income level and the current income at a given shift of a taxi driver towards the hazard of stopping. This method faces the difficulty of having to make an educated guess on the potential reference income. Usually some variation of the mean income, including some stochastic component is chosen (Farber, 2008; Crawford and Meng, 2011; Martin, 2017). The evidence so far has been inconclusive with most studies finding mixed results that do not directly point to the direction of either neoclassical or reference-dependent preferences. Most recently Farber (2015) has found evidence pointing towards a positive wage elasticity, with only inexperienced drivers behaving as reference earners. Martin (2017) applied what was originally the basis for reference dependence to the subject, namely the prospect theory. Thus far the literature had assumed loss-aversion for the reference dependence, where the main driver of utility is whether the reference point is reached or not. The prospect theory differs from loss-aversion in that the relative difference between current income and reference point is crucial towards evaluating marginal utility at any given point, with marginal utility being higher around the reference level.

The minibus sector differs in several features compared to the traditional taxi sector, which leads to different incentives in driving behaviour. It can be observed that fares per passenger fall into a very small range of 6–12 ZAR in Rustenburg and 9–15 ZAR in Cape Town. The fare is generally set in an area by that area's association, and so fares will change as a traveller moves from one association to another. Consideration is taken of consultations with passengers' economic ability to pay. Fares do not surge and generally only increase every 2 or 3 years (not annually). Hence, there seems to be some form of mutually accepted prices, but it is not immediately clear what separates a 15 ZAR trip from a 9 ZAR trip. The fare is loosely related to distance, so a longer distance trip will generally cost marginally more than a shorter distance trip. However, based on the flat

nature of this pricing scheme (see Neumann et al., 2015), strong incentives exist to increase the number rather than the length of trips during a shift. Therefore, the drivers are encouraged to end trips fast, as there are only marginal revenues to be gained from slightly higher fares for longer trips. The market is self-regulated by competition along a route amongst parties that share the monopoly for that route. The most striking difference however is the fact that minibus taxis have an official load capacity of 15 passengers (though at times more passengers are fitted in), hence they are able to pick up multiple passengers as they complete their trip. On the contrary, a classical taxi is marked as occupied once a passenger (or a group of passengers) has entered the vehicle. This completely alters the structure of incentives for a South African minibus taxi driver, since he or she can search for additional passengers before alighting the passengers at their destination along the same route.

Due to this feature, instead of evaluating a reference level at a daily level it might be just a single trip that has to reach a certain amount of income in the mind of a taxi driver for him to cease to look for more passengers and to proceed to the destination for that trip. It is conceivable that drivers in South Africa follow a two-tiered model of reference-preferences, one for each trip and one for each day (or possibly week). In the following a utility function for a taxi driver for taking another trip and looking for another passenger is modelled. Assume the following revenues Y_{ds} for driver d on shift (or day) s for trip t ,

$$Y_{ds} = \sum_{t=1}^T Y_{dst} \quad (1)$$

and his income for any trip be defined as:

$$Y_{dst} = \sum_{n=1}^N P_{dstn}, \quad (2)$$

where n is defined as the number of passengers on trip t and P as the revenue per passenger. Hence

$$Y_{ds} = \sum_{t=1}^T \sum_{n=1}^N P_{dstn}. \quad (3)$$

Based on Crawford and Meng (2011) the utility function of a driver is additively separable into a consumption part:

$$U_1(Y), \quad (3a)$$

$$U_2(H), \quad (3b)$$

where the utility U_1 is an increasing and concave function derived from earned income (Y), meaning that from $Y_1 \leq Y_2$ follows $U_1(Y_1) \leq U_1(Y_2)$ and U_2 is a decreasing and concave function dependent on hours worked (H), i.e. from $H_1 \leq H_2$ follows $U_2(H_1) \geq U_2(H_2)$.

The reference-dependence part:

$$R(Y, H | Y^*, H^*) \quad (3c)$$

indicates the gain-loss utility with Y^* as reference level income and reference working hours, H^* , meaning the targeted daily income and hours worked. This leads to a driver's total utility

$$V(Y, H | Y^*, H^*) = (1 - \mu)(U_1(Y) + U_2(H)) + \mu(R(Y, H | Y^*, H^*)) \quad (4)$$

where μ is a constant parameter with $0 \leq \mu \leq 1$, Y can be interpreted as any of the incomes in Eqs. (1)–(3), H is the time already worked, the asterisk marks the reference level of a driver.

Note that

$$\frac{\delta U_1}{\delta Y} > 0, \frac{\delta U_2}{\delta H} < 0. \quad (5)$$

The utility from earning income is positive and from working more hours negative. Both U_1 and U_2 can be thought of as being evaluated over any timeframe, be it a day or even just a single trip. The marginal utility ‘kinks’ in income and hours worked at their respective reference levels Y^* and H^* . Marginal utility before (after) reference level is higher due to the constant factors than after reference point for income (hours). Thus, the function is continuous but not linear, or only partwise linear. The utility derived from deviating from the reference level is defined as follows:

$$\begin{aligned} R(Y, H | Y^*, H^*) = & I[Y > Y^*] (U_1(Y) - U_1(Y^*)) \\ & + I[Y \leq Y^*] \gamma (U_1(Y) - U_1(Y^*)) \\ & + I[H > H^*] \rho (U_2(H) - U_2(H^*)) \\ & + I[H \leq H^*] (U_2(H) - U_2(H^*)), \end{aligned} \quad (6)$$

where I is an index function that evaluates to 1 if the condition is fulfilled, γ is a constant factor stating the disutility of not having reached reference level income and ρ the disutility of having worked more time than reference level. Both parameters should be ≥ 1 as this would indicate loss-aversion opposed to gains-aversion in case of < 1 which would be entirely irrational. This parametrization also rules out a reference dependence based on prospect theory as constant factorization does not account for the increasing (decreasing) marginal utility for the loss (gains) region.

Any empirical implementation of the aforementioned theoretical framework for our dataset has to take into account some limitations and differences of minibuses taxis. Implementing an instrumental variable for Y^* and H^* on a per driver basis through averages and an unobserved random error will almost certainly not yield stable results as the sample includes around 427 drivers with just about 18 trips on average, hence there will be too little variation in revenue per driver. As Farber (2008) has shown, working with a common reference level across all drivers is problematic. Hence a shift-level reference dependence is hard to research through this method with the given data set. However, it is possible to estimate the increase and decrease in stopping hazard for the shift of any driver through splines of achieved earnings and already worked hours comparable to Farber (2005) or Martin (2017).

As this paper aims primarily to find out about driver behaviour in dependence of the hours and earnings variables, the remainder of this paper is concerned with finding reasonable estimates for H^* and Y^* with using the method of interpreting increasing hazards of stopping as indication of a possible reference income/hours worked. To this end, due to the non-linear nature of reference-dependent preferences and their respective utility functions, estimation methods that can reflect non-linear relationships have to be employed. For example, one could expect a sharp increase in the hazard of stopping once some reference level is reached. This would mean a sharp increase of the dependent variable after some value of the independent variable is reached, indicating a distinctively non-linear relationship. Therefore, a spline regression has been chosen as empirical model to account for possible nonlinearities. A spline regression avoids the shortfall of pure polynomial approaches when modelling a non-negative variable without a positive limit (e.g. revenues or hours worked) by dividing a variable into a set of a certain number of splines, each represented by a bounded category with no intersections between them. An extensive derivation and reasoning can be found in Appendix A. After

analyzing the parametric regressions, generalised additive models (GAM) are also applied to avoid the rigid modelling approaches that standard regressions offer when exploring non-linear relationships. In addition, the practice of somewhat arbitrarily choosing the ranges of a spline is replaced by a data-driven approach. Fitting a LOESS model severely limited the fit of the envisaged model, for that reason they are not included in this paper.

Results

Data description. A multi-phase municipal public transport survey was undertaken in the city of Rustenburg to collect public transport data on operations of minibuses taxi associations running their services within the municipal boundary. The surveys, which took place on 49 days between November 2016 and September 2017, were onboard observation surveys and were executed by means of the GoMetro Pro mapping application, a mobile application that enables field collectors to quickly and easily map out transport routes. The methodology consisted of cordon counts coupled to onboard surveys. Operational data (7765 observations or trips) was captured over a full operating day by a representative sample of vehicles of the larger population on predetermined days and periods of the month that were selected in collaboration with the taxi associations to accurately reflect the monthly operations. The sampling was done on a $\geq 80\%$ confidence level with a 20% confidence interval when sampling a fleet of a homogenous population without stratification.

The data was collected as part of a survey that was undertaken for the implementation of the Rustenburg Integrated Rapid Public Transport Network (IRPTN). The purpose of the survey was to present operations analysis results that would be used for spatial and business planning of IRPTN and the minibuses taxi industry. An additional dataset (4012 observations) was also collected in metropolitan Cape Town with a number of associations across the City. These associations were sampled on 27 days between April and May 2017 on a trip by trip basis—and not specifically profiling an individual driver for the same daily duration as the Rustenburg dataset.

Revenue data and operational data per vehicle for a full day of operations was collected by the mobile application, repeated on selected days in a month. The data was inferred by post-processing of the data and applied to individual business models for each sampled minibuses taxi. The individual business models were then extrapolated to the population as a representative sample of the population to determine the system’s revenue of an association.

Engaging with the taxi industry was one of the most critical parts of planning the survey. This engagement involves comprehensively informing the executive committee of the minibuses taxi association about the survey, i.e. why the survey is being conducted and what data will be collected. The trust built during the initial engagements with the associations is important for ensuring that the surveys proceed with full cooperation from daily operational staff of the association. The taxi industries have a deep understanding of their operations and, therefore, are key stakeholders in successfully planning out the surveys in order to portray an accurate picture of the industry operations. Once agreements were reached, trust was earned, and the taxi associations cooperated with the survey team. They assisted the survey team in gathering insights about what to expect during the survey and were available to assist the survey team overcome challenges in conducting the surveys.

Mappers boarded the allocated minibuses taxis with a mobile device equipped with the GoMetro Pro App. Operations of the sampled minibuses taxis were surveyed for the entire day collecting

data on working hours, number of passengers boarding and alighting as well as respective revenues. Socio-economic data on passengers was collected but not used in this analysis due to missing observations. Separate trips were captured and logged for the both the forward and return trips of each individual route, i.e. one trip from the CBD taxi rank to Tlhabane and one trip back from Tlhabane to the CBD taxi rank. Vehicles being surveyed operated from the early morning to the evening. Two mappers captured a vehicle's operations in one day by working two shifts, one in the morning and the other in the evening. The mappers changing shifts would communicate to determine the handover location. This was then communicated to the project managers who would confirm and keep record of the devices being handed over. Due to cooperation between the managers, supervisors and mappers, no devices were stolen or damaged during the execution of this survey.

Lastly, the data have been cleaned from observations that obviously expose false values: 1. Trips whose average speed is lower than 10 km/h, an average speed that is too low even when taking into account sustained traffic jams. 2. Trips with an average speed of over 80 km/h. Considering that Rustenburg is a medium-sized town with few possibilities to drive on highways or fast lanes, these observations are likely to result from recording errors.

Model description. The employed empirical models in our analysis are parametrical (polynomials and splines) and non-parametrical (generalised additive models) models. We regress splines as defined above (and in Appendix A) on a binary variable indicating whether the observed trip is the last trip of the day for the driver by varying methods as introduced in the last section. Splines are introduced for work hours and revenues, cumulated after every single trip on a daily basis, meaning how much has a driver earned and worked on the current day after each trip on that day. Work hours have been defined as the difference between the time when a given trip ends and the start time of the first trip of a day of a driver. It must be noted that for some observations work hours tend to be very long, as there are large time gaps between trips.

Unfortunately, the sample supplies no information on whether the driver actually spends the time in between waiting for passengers or whether he takes a break. Further interrogation of the driver situation by observation revealed that drivers are either waiting in the vehicle queue to load passengers (a wait that can exceed one hour) or are taking a break to clean their vehicle, rest or eat. Alternatively, the cumulated minutes actually spent driving have been added, but ultimately showed very little effect and significance. The dependant variable is defined as 0 for all trips except if that trip is the last trip of that taxi and driver for the particular day, then it assumes 1. The splines for revenues and hours worked have been set as the borders of the inner four quintiles for revenues and hours, respectively (meaning 20%, 40%, 60% and 80%) in Table A.2 (see Appendix B for regression tables). Respectively in Table A.3, quintiles of revenues and hours for the observations that constitute a last trip (hence $y = 1$) have been set as boundaries to give weight to revenues levels that actually constitute a 'last trip'. A set of control variables have been added to the model, which include time of day after the trip has ended, day of week variables and a set of driver fixed effects. The models as specified above relate to the columns in Tables A.1–A.3, with the first column always being a simple ordinary least-squares (OLS) regression for reference, and the second and third columns representing binary regressions with logit and probit link functions respectively (see Appendix A for reasoning).

Parametric regression results. In Table A.1 (Appendix B) the regression table for the simple model with polynomials is depicted. Highly significant and positive results on the second and third polynomial as well as negative values on the first polynomial direct towards a non-linear relationship between revenues and the probability of stopping. However, due to the numerous shortcomings of simple polynomial models, e.g. their unreliability on large scales such as revenues for a currency like Rand, Table A.1 can only serve so far as to point us towards nonlinearity. No significance can be attributed to the estimators on the number of hours worked, notably, after the time of day and the end of a trip has been controlled for. The estimators on the time of day binary variables are highly significant and positive especially for the evening hours starting from 5 p.m. but have not been reported in the tables for ease of reading. Ultimately, they point towards a certain period in the evening when drivers stop.

The spline regressions deliver an equally mixed picture, with the simple quintiles (without prior subsection for last trip values), delivering the most significant results, whereas Table A.3 does not show any significance in almost any spline. Although both selected sets of splines result in somewhat reasonable values, almost all variables are not statistically significant even at the 10 % level. The notable result is that especially the splines in the middle in Table A.2 show a negative coefficient just as the estimator on the simple count of accumulated revenues, again confirming the existence of non-linearities, but failing to reach significance even at higher levels. Additionally, two of the splines on the estimators on hours worked picked up significance, indicating a subsiding likelihood to stop in the first hours of a working day.

In summary, parametric models have shown so far that the minibus taxi drivers chose the end of their shift primarily based on what time it is rather than how much money they have earned or how much they have worked already, as the much larger estimates on the hours worked splines and the not reported results on the hour of day effects prove (note however the definition of hours worked due to the sampling). The results with respect to revenues are inconclusive due to their significance but show non-linear patterns. Including additional splines and polynomials would smoothen the curve but have not delivered tangible results.

GAM results. As a possible cure to the inconclusive results on revenues, GAM are applied in this section: A non-parametric regression that avoids the rigid modelling approaches that standard regressions offer to explore nonlinear relationships. In addition, the practice of somewhat arbitrarily choosing the ranges of a spline is replaced by a data-driven approach.

The usual set of control variables for day, time of day and driver fixed effects have been parametrically included in the model, whereas the variables of interest, accumulated revenues and hours worked have been chosen to be included in the model in a functional form. For the link function, the logit model (see Fahrmeir et al., 2013) has been chosen as it persistently represented the best fit in the parametric models, measured through Akaike information criterion (AIC) values. The fitted log-likelihoods are shown in Figs. 1 and 2. The AIC for the former results in 4167 with both smooth terms highly significant, for the latter in 4101, suggesting a better fit for the model that still includes hourly fixed effect and being equally good fit for the model as the parametric regressions in Tables A.1–A.3. *Hours worked* loses its significance entirely once the hourly effects are included (see Tables A.4 and A.5 in Appendix B for statistical inference). Figure 1 shows the relationship when hourly fixed effects are not included.

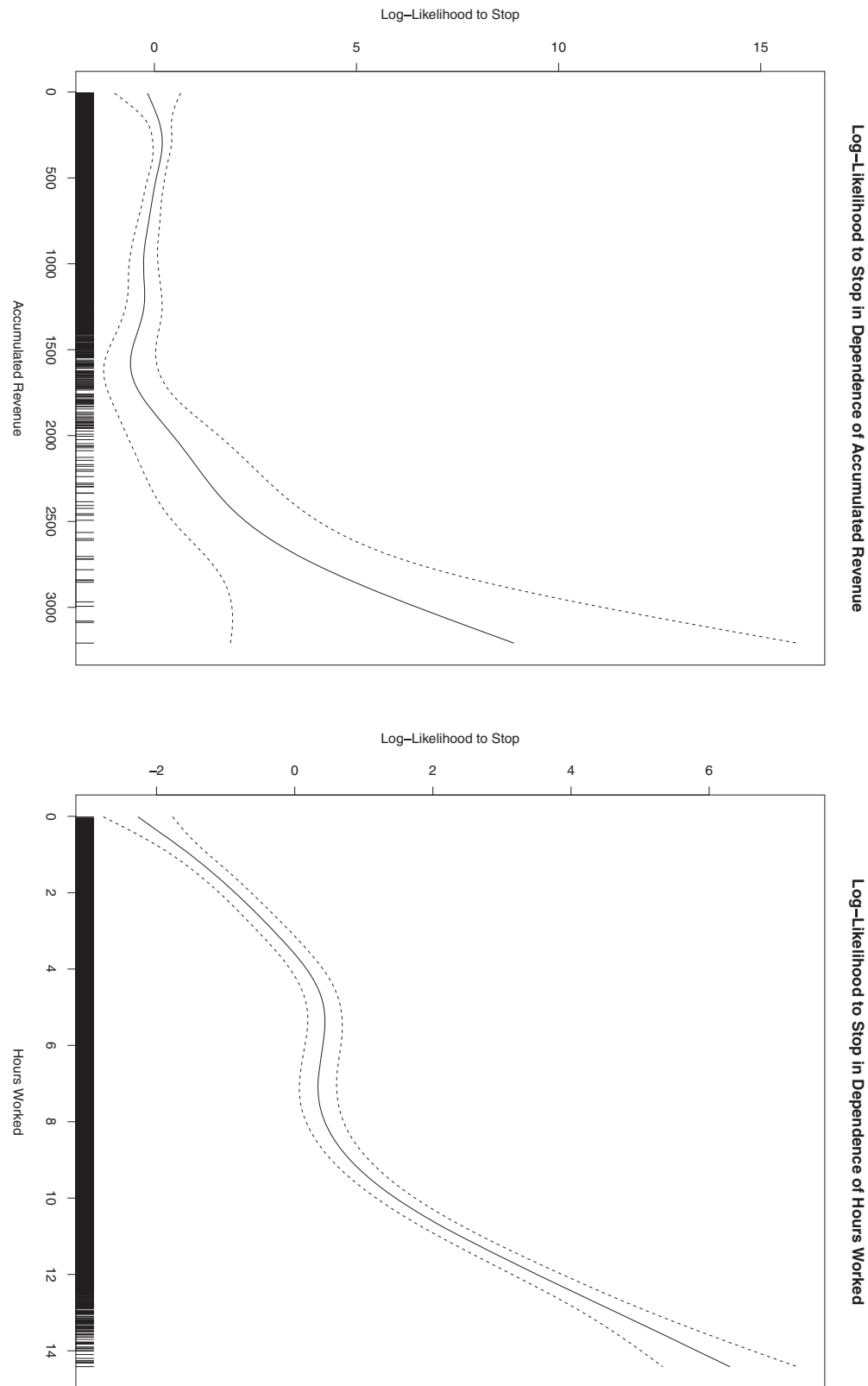


Fig. 1 Line indicating the log-likelihood to stop obtained by a regression with a generalised additive model including all parameters, with the exception of revenues accumulated and hours worked which were chosen as variables to smooth over. No hour fixed effects were included. 95% confidence interval included. Smoothing parameter of 2.47 chosen for accumulated revenues and 5.51 for hours worked. See more particular regression results for smooth terms in Table A.4. Source: Own representation, Data: GoMetro (2017).

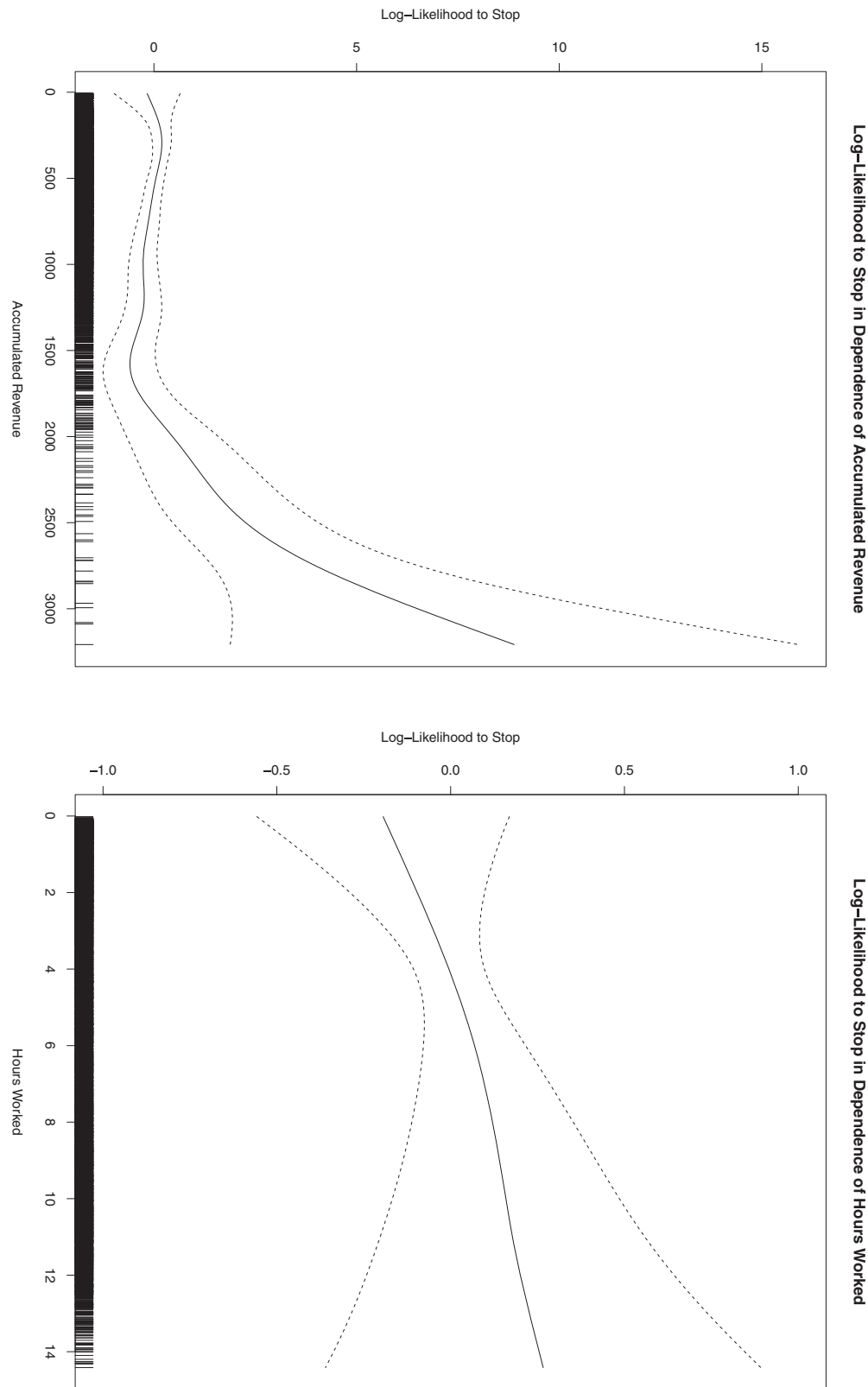


Fig. 2 Line indicating the log-likelihood to stop obtained by a regression with a generalised additive model including all parameters, with the exception of revenues accumulated and hours worked which were chosen as variables to smooth over. Hour fixed effects were included. 95% confidence interval included. Smoothing parameter of 7.59 chosen for accumulated revenues and 1.322 for hours worked. See more particular regression results for smooth terms in Table A.5. Source: Own representation, Data: GoMetro (2017).

As is visible in Fig. 2 no changes occur to the smooth term of revenues once hourly effects are included. This once again confirms the finding of that revenues at a day level have, if at all, only a very minor influence on the likelihood or hazard of stopping for South African minibus taxi drivers. We can only assume that having had extraordinarily high earnings significantly impacts the likelihood to stop, but the limited number of observations may make this claim spurious.

To the contrary, the number of hours worked seems to have a high impact on the probability to stop, at least in Fig. 1. In addition, the GAM model has revealed an interesting non-linear relationship that the parametric model seems to have missed. For the first four hours of work on a day a negative impact on the likelihood to stop can be observed, which is immediately intuitive. In the range from 5 to about 8 h the likelihood stays constant at about 0.5 (probability of around 0.6). After about 9 h the likelihood rapidly increases. This pattern could be explained by the potential existence of two distinct types of taxi drivers: One more professionally involved, working a full day, probably being part of a rather professional company that owns a number of buses, which then probably account for the drivers who start to stop after 9 h. The others being drivers who pursue the profession in a more loose manner, potentially as a side job to earn an additional income and thus only have time for a few hours of driving on a given day, thus explaining the relatively high probability to stop in the dataset after 5 h, as they might have less time to pursue the driving occupation for longer on a given day. This hypothesis could be tested by inventing some metric to tell these two (or more) types apart and analyse their behaviour separately, which is not possible with the current dataset as no such information is provided.

Notably, as is evident in Fig. 2, this relationship entirely breaks down once hourly fixed effects are included. Only a very small effect is visible from a -0.2 to 0.2 in log terms with very wide confidence intervals, practically indicating that there is no relationship. The hourly fixed effects expose a strongly increasing trend the later it is in the day (see Table A.6). The surprising non-relevance of accumulated revenues for the likelihood to stop before a very high threshold of about 1500 Rand, as well as the irrelevance of hours worked once the respective fixed effects are taken into account can have two explanations: Firstly, taxi drivers mostly 'work hours' instead of revenues and do not expose the behaviour of self-employed taxi drivers elsewhere, e.g. in New York City. That is, they decide on the end of their shift based on the current time rather than Rand earned, basically exposing the behaviour of ordinary employees, starting in the morning and finishing in the evening. If the fixed effects are considered (e.g. higher probabilities of stopping the further the day advances) it can be assumed that most taxis end in the evening hours and those that start later also work shorter hours. Another hypothesis is that a given day is not the relevant bracket of evaluation for the drivers and it could rather be a week or a month or even just a single trip.

Discussion

According to the theory of reference levels (or the prospect theory) drivers should base their decision to work (or to cease working) on what they have already attained rather than what they expect they can potentially gain in terms of worked time and revenues earned. Thus, if drivers were perfect reference earners they would almost immediately stop working when their relevant reference level is reached, be it on a day level or a trip level.¹ This should be visible in the models as a rapid increase in the probability to stop, meaning either of the variables drawn up in this section. This conclusion stems from the theory which predicts

increasing utility before a reference level is reached and constant or almost constant utility thereafter, or vice versa if the variable in question yields negative utility (e.g. as working time). The class of GAM yielded an easy to interpret model that allows to draw conclusions on the applicability of the theory of reference dependence. Looking at Figs. 1 and 3 it can be concluded that drivers do not choose to end their working day based on how much income they have earned during the day but certainly rather based on how long they have worked and what time it currently is. It is evident that the probability to stop drastically increases in the time frame from 10 to 12 h worked, whereas in that particular time frame they tend to stop later the more income they have earned already. This suggests that although drivers certainly work a somewhat fixed number of hours per day and usually stop in the evening hours, there appears to be some flexibility to react to the market situation. If we assume that a high-income level attained signals a 'good day' then it is likely that drivers put up an extra hour to reap the benefits of increased demand (or low supply). This holds true, vice versa, for very low income levels attained (>300 akin to the first quintile of income earned on the last trip of the day) where a comparatively high probability to stop of 0.3 after 4–6 h already, indicating that drivers stop sooner on 'bad days'. We can also see in Fig. 3 that the probability to stop reverses and diminishes from 6–8 h worked on especially so for income levels >400 , however residing at very low levels (<0.3) across all income levels until the already mentioned 10–12 h of worked time.

From Fig. 2 paired with Fig. 3 we can comfortably conclude that drivers chose the length of their working days primarily based on a fixed work schedule, usually choosing to stop in the evening hours, whereas the starting time can be different across drivers. For this reason, Fig. 3 shows a higher likelihood to stop after 4–6 h. This relationship is due to drivers that start around midday and end their shift at night too, after having worked only 6 h. Nevertheless, most drivers will work a full 10–12 h from morning until evening. The inclusion of hourly fixed effects in Fig. 2 shows that these behavioural patterns are the underlying reason for the stopping decision on a given day. Neither accumulated revenues (p -value = 0.04, but with hardly any effect until a large amount of revenues) nor hours worked (p -value = 0.47) seem to really matter once the hourly effects are considered. This is probably also due to the fact that high revenues, many hours worked, and an evening hour will almost always fall together. It can be concluded that time is the relevant variable a driver looks at when deciding whether to add another trip or not.

To summarise, it can be stated that minibus taxi drivers do indeed behave according to a certain reference level, namely, 10–12 h a day (which mostly translates to simply stopping in the evening). Nevertheless, they certainly also expose profit-maximizing behaviour, as they shorten their working day/trip if the conditions are unfavourable and lengthen them when the conditions seem profitable.

As Sheldon (2016) indicates, new drivers in the market may actually show signs of reference earners (which he refers to as income-targeting), however, due to a learning process, drivers eventually show profit-maximizing behaviour. His reasoning is that new drivers adopt to a mentality to avoid procrastinating and underperformance, but soon intertemporal substitution allows for better returns as more income can be achieved during favourable conditions and leisure can be expanded through the higher income. This clearly implies neoclassical behaviour as described in sections "State of research" and "Theoretical framework", and contradicts the findings by Cramer et al. (1997) who identified negative labour supply elasticities for taxi drivers, setting an influential mark in behavioural economics research. Nevertheless, reference-dependent preferences are a valuable framework for

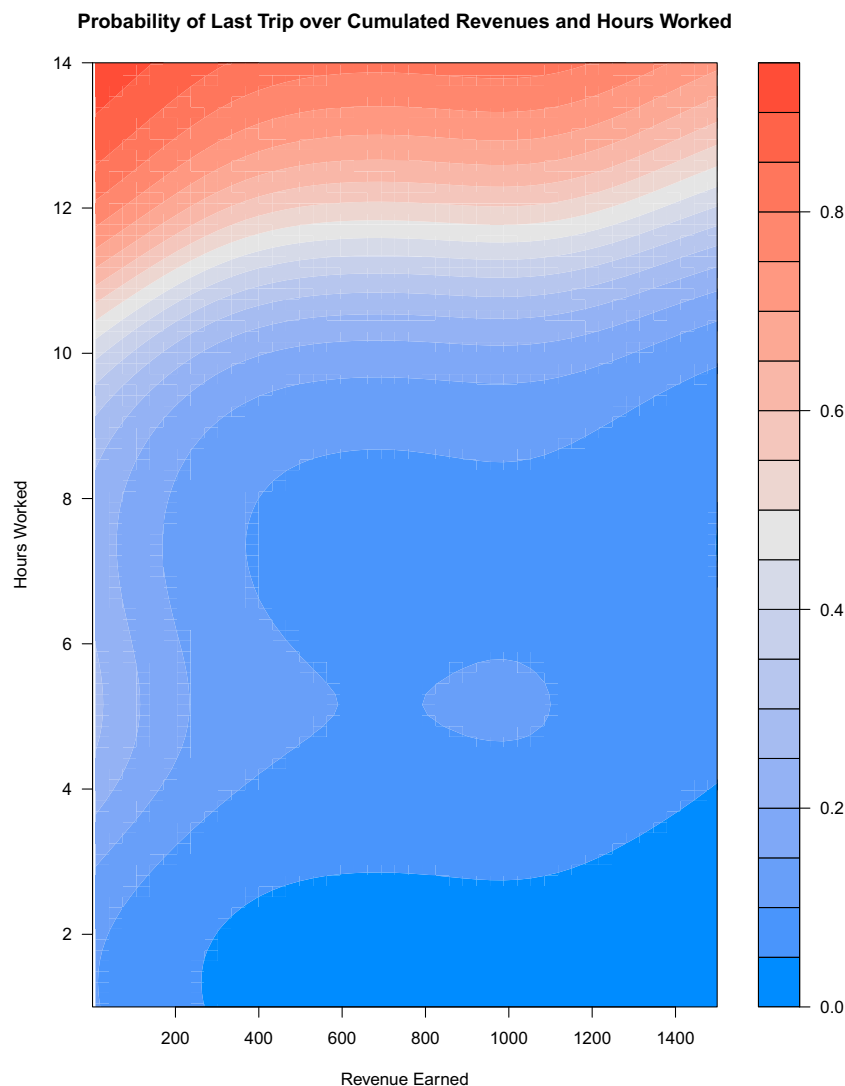


Fig. 3 Heatmap indicating predicted probability of ending the shift in dependence of hours worked and revenues earned. Results were obtained from a strictly non-parametric regression according to the GAM method introduced, but just the time and revenue variable used as predictors (thus excluding all fixed effects), whereas both were chosen to be smoothed over. Range of data has been chosen based on the number of data points 7.5. E.g. too few data points fall after 1500 ZAR earned. Source: Own representation, Data: GoMetro (2017).

behavioural analysis, especially when reference points are clearly defined, e.g. in stated choice experiments. Moreover, it has to be acknowledged that reference points are a dynamic concept: driver behaviour may vary over time and generations, e.g. when drivers follow a learning process or when generations shift and new paradigms become important. As Börjesson and Eliasson (2014) state “A couple of years into the future, traffic on a road may look similar as today—but it will not be the same drivers”. However, they also highlight that reference dependence, despite its intuitive appeal, might be a phenomenon that vanishes over time with experience which follows the reasoning by Sheldon (2016) and Martin (2017). Consequently, reference dependence may be applicable for short-term valuation and behavioural analyses (e.g. it has been applied to congestion pricing) but infeasible for long-run assessments.

The model and data envisaged in this paper have some shortcomings. Firstly, the way that the *hours worked* variable has been formulated has the drawback that it cannot be differentiated as to what the driver did when he was not driving, whether he was taking a break or was waiting in line or else. As the data did not yield any such information the notation had to be chosen that

way. Other empirical implementations (e.g. Farber, 2005) showed that this information can be important. Secondly, many more binary variables could have increased the explanatory power of the model, e.g. start and/or stop area fixed effects, weather and others are conceivable, see for example Martin (2017) or Farber (2015). Particularly, area affects could have had some impact, but the meaningful sub-setting of the entire Rustenburg area into different areas was beyond the scope of this paper. In line with Martin (2017) any more information about the driver, e.g. socio-economic and certainly experience, would add a highly interesting layer to the results. Sheldon (2016) provides a rare analysis on how learning (working experience) may affect taxi driver behaviour. Data of this nature were not included in the data set. The model as it has been chosen is additive in nature, this was chosen to make the effects of the revenue and time variable separately visible. It could well be that time and revenue affect driver behaviour in an interactive form. This possibility has been spared out to retain individual interpretability of the variables. This paper has suggested day and single trips as relevant brackets for the drivers to base their decisions on, but it is also conceivable that drivers look foremost at their income on a weekly or monthly

level, e.g. the same interval in which they turn over the revenues to the owner. The survey method of the data again precludes this area of research, as drivers were only followed consistently through one individual day, hence the data does not yield enough information on these time brackets.

Conclusion

The theoretical underpinning of the empirical model of this paper has been reference-dependent preferences, which would predict behaviour opposed to sheer profit maximization (e.g. driving more on 'good days' and less on 'bad days'): After a certain threshold in hours or revenue is reached stopping probability should drastically increase. A multitude of models has been tested to possibly define the point of reference if it existed. Potential reference points on a trip and stop level have been accounted for. A model of linear splines, local regression and GAM have been applied to the problems, with the latter consistently showing the most significant (according to AIC and log-likelihood) and reasonable results in both cases. The results have shown that on a trip level, drivers clearly aim to work 10–12 h until the evening hours with about an hour of flexibility if the accumulated profits are low (less work) or high (more work) pointing towards some profit maximization part in their behaviour. The shortcomings of the envisaged approach have been lined out in the last section and it is clear that more tailored data is needed to define more precise results. Nevertheless, this paper has, in a first instance, established that South African minibus taxi drivers cannot clearly be attributed to either reference dependence or profit maximization behaviour. They indeed seem to expose both behaviours to some degree, with reference dependence being more spelled out, as certain thresholds were indeed pinpointed. These findings provide a fruitful base for paratransit optimization research in other developing countries with similar informally organised transport modes. Additional research in the field of psychology could complement our data-driven empirical approaches.

This research holds value for future organizational attempts directed at the minibus taxi sector or global paratransit services alike. The total absence of technology in the conduct of this business is remarkable and holds great potential for gains in efficiency and effectiveness. Especially so since the technology is probably soon to be available and the usage of mobile devices has proliferated in South Africa as elsewhere. A technology-based approach to modern mobility in developing countries is likely to be reliant on automated algorithms to assign and navigate minibus taxis and other paratransit. Therefore, insights like these on minibus taxi driver behaviour can be used to set heuristics in respective algorithms to predict and, thus, balance out mobility supply throughout a day. As to what extent machine learning may benefit from these findings is left for further research. In addition, potential efficiency gains are likely to imply beneficial indirect effects, e.g. reduction of emissions from paratransit as indicated by Coetzee et al. (2019). Notwithstanding, environmental effects of optimised paratransit in developing countries requires further intensive research.

Data availability

The datasets analysed during the current study are not publicly available due confidential company data by GoMetro but are available from GoMetro (justin@gometroapp.com) on reasonable request.

Received: 8 February 2020; Accepted: 15 May 2020;

Published online: 22 June 2020

Note

- 1 Note: We conducted an additional analysis to assess whether there is an additional layer of reference income for minibus taxi drivers. Using the same models as before we used intra-trip behaviour, where the drivers decide primarily based on how many passengers have so far boarded during a trip when and where to 'end' it and to proceed to the destination area. While day-level earnings (results presented in this paper) have no significant effect on the stopping rate up until a very high level, the level of income closely proxied through the number of passengers boarded has a highly significant and non-linear influence on how drivers conduct their trips.

References

- Barrett J (2003) Organizing in the informal economy: a case study of the minibus taxi industry in South Africa. No. 993581583402676. International Labour Organization
- Börjesson M, Eliasson J (2014) Experiences from the Swedish Value of Time study. *Transportation Res Part A* 59:144–158. <https://doi.org/10.1016/j.tra.2013.10.022>
- Brake J, Nelson J, Wright S (2004) Demand responsive transport: towards the emergence of a new market segment. *J Transp Geogr* 12:323–337
- Braumann A, Haferburg C (2010) Fußball-WM 2010 in Südafrika–Platzverweis für die Minitaxis? In: Haferburg Christoph, Steinbrink Malte (eds) *Megaevent und Stadtentwicklung im globalen Süden. Die Fußballweltmeisterschaft. Brandes & Apse, Frankfurt a. M.*, pp. 166–181
- Buja A, Hastie T, Tibshirani R (1986) Linear smoothers and additive models. *Ann Stat* 17(2):453–510
- Camerer C, Babcock L, Loewenstein G, Thaler R (1997) Labor supply of New York City cabdrivers: one day at a time. *Q J Econ* 112(2):407–441
- Cervero R, Golub A (2007) Informal transport: a global perspective. *Transp Policy* 14:445–457
- Chou Y (2000) Testing alternative models of labor supply evidence from taxi-drivers in Singapore. Department of Economics Research Paper, 768. University of Melbourne, pp. 1–39
- Coetzee J, Zhuwaki N, Blagus D (2019) Demand-responsive transit design methods and applications for minibus-taxi hybrid models in South Africa. In: *Southern African Transport Conference*, 2019. <https://repository.up.ac.za/handle/2263/74245>
- Crawford V, Meng J (2011) New York City Cab Drivers' Labor Supply Revisited: reference-dependent preferences with rational-expectations targets for hours and income. *Am Econ Rev* 101(5):1912–1932
- Del Mistro R, Behrens R (2015) Integrating the informal with the formal: an estimation of the impacts of a shift from paratransit line-haul to feeder service provision in Cape Town. *Case Stud Transp Policy* 3(2):271–277
- Fahrmeir L, Kneib T, Lang S, Marx B (2013) *Regression—models, methods and applications*. Springer, Berlin, Heidelberg
- Farber H (2005) Is Tomorrow another day? The labor Supply of New York City Cabdrivers. *J Political Econ* 113(1):46–82
- Farber H (2008) Reference-dependent preferences and labor supply: the case of New York City Taxi Drivers. *Am Econ Rev* 98(3):1069–1082
- Farber H (2015) Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *Q J Econ* 130(4):1975–2026. <https://doi.org/10.1093/qje/qjv026>
- Fernández J, Cea JD, Malbran R (2007) Demand responsive urban public transport system design: methodology and application. *Transp Res Part A* 42(7):951–972
- Germinah E, Chiloane T (2016) Factors affecting the implementation of the taxi Recapitalisation Project: the Department of Transport. *Probl Perspect Manag* 14(4):25–32
- GoMetro (2017) Records on minibus taxi trips in Rustenburg and Cape Town, South Africa in 2016 and 2017 (not publicly available)
- Hastie T, Tibshirani R (1986) Generalized additive models. *Stat Sci* 1(3):297–310
- Jokinen J-P, Sihvola T, Hyytiä E, Sulonen R (2001) Why urban mass demand responsive transport? IEEE Forum on integrated and sustainable transportation systems, Vienna, Austria, June 29–July 1, pp. 1–6
- Joubert J (2013) Gauteng: paratransit—perpetual pain or potent potential? In: *Institute for Mobility Research (eds) Megacity mobility culture*. Springer, Berlin, Heidelberg, pp. 107–126
- Kahneman D, Knetch J, Thaler R (1991) Anomalies: the endowment effect, loss aversion, and status quo bias. *J Econ Perspect* 5(1):193–206
- Kőszegi B, Rabin M (2006) A model of reference-dependent preferences. *Q J Econ* 121(4):1133–1165
- Martin V (2017) When to quit: narrow bracketing and reference dependence in taxi drivers. *J Econ Behav Organ* 144:166–187
- McCaul C, Ntulim S (2011) Negotiating the deal to enable the first Rea Vaya bus operating company: agreements, experience and lessons. In: *Proceedings: 30th Southern African transport conference*, Pretoria, South Africa
- Mossay P, Picard P (2019) Spatial segregation and urban structure. *J Reg Sci* 59(3):480–507
- Neumann A, Röder D, Joubert JW (2015) Toward a simulation of minibuses in South Africa. *J Transp Land Use* 8(1):137–154

- Nguyen-Hoang P, Yeung R (2010) What is paratransit worth? *Transp Res Part A* 44(10):841–853
- NHTS (2013) Statistics South Africa. National Household Travel Survey
- Phun V, Yai T (2016) State of the art paratransit literatures in Asian developing countries. *Asian Transp Stud* 4(1):57–77
- Schalekamp H, Behrens R (2009) An International Review of Paratransit Regulation and Integration Experiences: lessons for Public Transport System Rationalisation and Improvement in South African Cities. In: *Proceedings of the 28th Southern African Transport Conference*, pp. 442–550. <https://repository.up.ac.za/handle/2263/11968>
- Schalekamp H, Behrens R (2010) Engaging paratransit on public transport reform initiatives in South Africa: a critique of policy and an investigation of appropriate engagement approaches. *Res Transp Econ* 29(1):371–378
- Schmidt M-A (2018) The Daily Labor Supply Response to Worker-Specific Earnings Shocks. Working Paper
- Sheldon M (2016) Income targeting and the ridesharing market. Unpublished manuscript
- Steiger J (2009) An introduction to splines. <http://www.statpower.net/Content/313/Lecture%20Notes/Splines.pdf>. Accessed 19 Jul 2018
- Thakral N, Tô L (2017) Daily labor supply and adaptive reference points. Unpublished manuscript
- Thaler R (1980) Toward a positive theory of consumer choice. *J Econ Behav Organ* 1:39–60
- Tversky A, Kahneman D (1979) Prospect theory. An analysis of decision under risk. *Econometrica* 47(2):263–291
- Tversky A, Kahneman D (1991) Loss aversion in riskless choice: a reference-dependent model. *Q J Econ* 106(4):1039–1061
- Venter C (2013) The lurch towards formalisation: lessons from the implementation of BRT in Johannesburg, South Africa. *Res Transp Econ* 39(1):114–120
- Wood S (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J Am Stat Assoc* 99(467):673–686
- Wood S (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc (B)* 73(1):3–36
- Woolf S, Joubert J (2013) A people-centred view on paratransit in South Africa. *Cities* 35:284–293

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1057/s41599-020-0508-2>.

Correspondence and requests for materials should be addressed to J.S.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020