African Journal of Business and Economic Research (AJBER)
Published considtently since 2006
(Online) ISSN 1750-4562 (Print) ISSN 1750-4554

Indexed by SCOPUS, UGC CARE List, IBSS, EBSCO, ProQuest, ABDC, SAJE, COPERNICUS, ERIH PLUS, CABELL, Sabinet and J-Gate

Vol. 19, (Issue 1), March 2024

Pp31 – 53

Modelling immigrants' participation in undeclared work in Gauteng province, South Africa

DOI: https://doi.org/10.31920/1750-4562/2024/v19n1a2

Kabeya Clement Mulamba
School of Economics
College of Business and Economics
University of Johannesburg
Email: mkclement@gmail.com
https://orcid.org/0000-0001-6428-8814

Abstract

The work that resulted in this article employed a hierarchical logistic model to test whether immigrants are more likely than non-immigrants to participate in undeclared work in Gauteng province. The empirical results were consistent with the structuralist exclusion hypothesis; immigrants were 1.81 times more likely than non-immigrants to participate in undeclared work. The results also imply that societal factors prompt immigrants to participate in undeclared work to a greater extent than non-immigrants. These discoveries suggest that policymakers should tackle the underlying structural problems that drive immigrants towards engaging in undeclared work, and strive to improve their working circumstances and facilitate their assimilation into formal employment arenas. One strategy the government could adopt to promote formalisation is to streamline the business registration procedure. The government should also prioritise easing the process of qualification transferability, to facilitate the recruitment of skilled and documented immigrants for formal employment opportunities.

Keywords: *Immigrant, undeclared work, Gauteng, South Africa*

JEL Classification codes: *C01, J01*

Introduction

The main objective of this study was to empirically test and determine the degree to which international migrants (hereafter referred to as immigrants) participate in undeclared work in the province of Gauteng, South Africa. This objective was based on the structuralist exclusion hypothesis, also referred to as the marginalisation hypothesis. According to this hypothesis, individuals belonging to marginalised groups (such as immigrants) are more predisposed to participation in undeclared work than others (Williams & Round, 2010). This occurs due to the regulatory environment, labour market conditions and social context; and when immigrants are undocumented, they are excluded from participation in declared or formal employment. Consequently, immigrants have no choice but to participate in undeclared work to ensure their survival. In this study, 'undeclared work' refers to paid work in informal employment. This definition aligns with that of Schneider (2008), in that generally, informal employment is not declared to the authorities (e.g. for tax administration).

This study focused on South Africa because of the long-standing issue of immigration into the country, as noted in OECD/ILO (2018); and mainly on Gauteng, as it is the most urbanised and industrialised province in South Africa, and home to the three largest cities in the country: Johannesburg, the City of Ekurhuleni and Tshwane. Consequently Gauteng attracts migrants from other provinces, and immigrants from neighbouring countries, who seek better opportunities regardless of their skills (Theodore et al., 2017). The region also hosts many asylum-seekers and refugees, whose principal aim is to seek protection (Blaauw & Pretorius, 2022). As a result, immigrants as well as migrants from other provinces compete with local residents in the labour market. However, as Theodore et al. (2017) argued, South Africa's economy remains rigid; it does not create enough jobs, leading to persistently high unemployment and inequality. This reality means that most immigrants cannot realise their dream of better opportunities (Akintola & Akintola, 2015); neither are asylum-seekers and refugees able to participate in the formal economy. Ultimately, whether immigrants are in the country legally, are refugees or have essential skills, participation in undeclared work is the only option for their survival.

The immigration issue remains a matter of public debate in South Africa. For instance, although it is a function assigned to the national sphere of government, in the 2021 municipal elections immigration was strongly featured in some political parties' campaigns and manifestos (Lekabe, 2021). Negative sentiments are shared by sections of communities who suggest that immigrants are taking jobs and other economic opportunities from South Africans. Consequently, some areas in Gauteng with a high concentration of informal immigrant-run businesses (and other small businesses in the eatery industry that were perceived to be employing more immigrants than South Africans) have been targeted by political and other social movement groups in the province (Mafata, 2022). However, even though perceptions of the subject created in the political arena are negative and without empirical evidence, the OECD/ILO (2018) reported that immigration has had some positive effects on the economy. Insights from the OECD/ILO report observe that well-managed immigration could contribute to the development of the province and the country. Furthermore, recent work by the World Bank showed a positive relationship between immigration and local employment in South Africa (World Bank Group, 2018). This study found that an increase (decrease) of one percent in immigration is associated with a 0.2 percent increase (decrease) in local employment, all things being equal.

The rapid immigrant population growth in Gauteng does constitute a challenge to the provincial government and municipalities' provision of essential services and social infrastructure, which cater for all residents. However, it is also well established that immigration brings dividends; for instance, Valencia et al. (2020) reported on the long-run fiscal dividends provided to Colombia by Venezuelan immigrants, as the government is able to collect additional revenue resulting from the immigrant's activities. But for Gauteng province to harness immigration dividends, aspects of immigration – including the participation of immigrants in undeclared work – must first be investigated and understood from a scientific point of view. Appropriate policies and regulations can then be designed to address the issues identified and ensure that immigration becomes a development tool for the province. Although many studies have investigated the issue of immigration in South Africa, there has been no estimate of the extent of immigrants participation in undeclared work in the province.

Through empirical evidence, this study sheds light on the structuralist exclusion hypothesis, estimating the magnitude and significance of the relationship between immigrants and undeclared work in Gauteng. It is also worth noting that while undeclared work contributes to the economy, its participants are exposed to low earnings and other poor working conditions. For immigrants, this is often exacerbated by challenges related to their integration into society. If not addressed, this set of circumstances will probably worsen the existing inequalities in Gauteng; hence a better understanding of immigration and undeclared work is required to inform the policy choices that should be made by authorities in the province.

To achieve the objective set, this study used the Quality of Life (QoL) survey published by the Gauteng City-Region Observatory ([GCRO] 2018) as the sole data source. One feature of the QoL survey is that respondents are nested within electoral wards (referred to as 'wards'), which are geographical units in Gauteng province. This structure is described as 'hierarchical', because the data has more than one level of information. The hierarchical nature of such data makes it necessary to carefully consider the issue of group effect when modelling the relationship between immigrants and undeclared work.

Based on the above, it was necessary to adopt a hierarchical logistic regression as a suitable framework for empirical analysis. In this framework (elaborated on further in the Model Specification section below, regarding methodology), the QoL survey data was considered to have two levels of information. Respondent characteristics (e.g. immigration status and gender) were recorded as first-level information, while the ward in which a respondent resided was considered second-level information. Consequently, one cannot assume that undeclared work (the dependent variable) for residents within a ward is independently distributed. In other words, there is the possibility that undeclared work by residents of a ward will be auto-correlated, because of unobserved ward-wide factors that affect all respondents within the ward similarly.

One contribution of this article is that it is the first study to assess the relationship between immigrants and undeclared work using data covering the entire Gauteng province. The assessment was undertaken to test whether the structuralist exclusion hypothesis is valid regarding immigrants to the province. To achieve this objective, the study adopted a hierarchical logistic modelling approach. This approach was chosen to enable reflection on possible homogeneous behaviour within wards, also dictated by the data structure, and to produce consistent estimates on which it would be possible to rely to inform policy decisions.

This article is structured as follows: a review of previous studies on the topic is presented in the next section. This is followed by the model specification and data sections, and then a discussion of the empirical findings. Finally, certain concluding remarks are made.

Literature Review

Most previous studies in South Africa have examined either the relationship between immigrants and informal self-employment, or the working conditions of informal workers (Blaauw & Pretorius, 2022; Charman & Petersen, 2015; Crush et al., 2015; Foster et al., 2021; Kalitanyi, 2010; Pretorius & Blaauw, 2019; McEwen & Leiman, 2008; Steyn, 2018; Thompson, 2016). None of these works estimated the extent of immigrants participation in undeclared work in Gauteng, as defined in the present article. However, the studies mentioned do illustrate, among other factors, the difficult working conditions to which immigrants are exposed in the informal economy in general. For instance, Blaauw and Pretorius (2022) reported that car guards felt their working conditions made them vulnerable to exploitation. The findings from the studies mentioned also illustrate the need to explore other aspects of immigration and informal work, thus supporting this article's rationale for estimating the propensity of immigrants to participate in undeclared work. This call for further research must also be understood in the context that immigrants face more challenges than non-immigrants. Therefore, as stated in the article's introduction, a better understanding of the association between immigration and undeclared work is necessary to inform the policy decisions that must be taken by authorities in the province.

Undeclared work preoccupies many researchers, but with the growing and extensive literature often focusing on European countries (Beręsewicz & Nikulin, 2018; Gashi & Williams, 2019; Williams & Efendic, 2021; Williams & Horodnic, 2015b, 2015d). For instance, Williams and Horodnic (2015a) tested the marginalisation thesis, producing mixed results. Their findings confirm that in Nordic countries, some marginalised individuals are more likely to participate in undeclared work than individuals from non-marginalised groups. In contrast, other marginalised groups – including women – are less likely to participate in undeclared work. These findings (Williams & Horodnic, 2015a) are an indication that the structuralist hypothesis of exclusion is required to be understood along with its nuances. In this article, these nuances are justified through the diverging contexts emerging from various regions (i.e. countries). It is thus necessary to empirically test this hypothesis case by case; conclusions should not be generalised from the experiences of European countries, for instance. Thus, one contribution of this article is its testing of the structuralist hypothesis as it relates to immigrants participation in undeclared work in the Gauteng province of South Africa. To the best knowledge of the author, the research question investigated in this article is the first of its kind.

Model Specification

The rationale for specifying a hierarchical logistic model[1]

Since the research undertaken in this study was empirical, a quantitative analysis approach (through a hierarchical logistic model) was employed to determine the relationship between undeclared work and the status of immigrants in Gauteng. This section discusses the rationale that was adopted for the estimation procedure and the data used for the analysis. In this regard a hierarchical logistic model was specified, for two main reasons. First, this method takes advantage of the richness of the data obtained from the QoL survey. This dataset comprised a comprehensive survey, conducted in Gauteng, to collect information about the quality of life of residents (including immigrants). In addition, the targeted outcome or dependent variable was formulated based on the survey question to determine whether a respondent participates in undeclared work. This dependent variable is a binary indicator known to follow a Bernoulli distribution. It necessitates considering a logistic regression to estimate the relationship between the dependent variable and respondents' immigration status.

Second, and as stated in the literature review above, from a statistical point of view, one reason for applying the hierarchical logistic regression is the hierarchical structure of QoL data. It requires careful consideration of the assumption often made in a traditional regression regarding the independence of the dependent variable's distribution. This assumption is relaxed in a typical hierarchical regression by the researcher selecting a model that allows for the clustering or dependence of observations. Because clustering is included in the model, a hierarchical specification helps to disentangle and estimate the two sources of variability in the behaviour of the phenomenon under investigation (undeclared work, in

[1] A non-technical note in the form of the Annexure is attached to this article for readers without a strong econometrics background.

this case). In other words, because QoL survey respondents are clustered within wards, one can estimate the within-ward variability of the participation of individuals in undeclared work under the model discussed. It is also viewed as the variability of participation in undeclared work, regardless of where a respondent resides.

Moreover, it is possible to estimate the between-wards variability of the phenomenon of interest using a hierarchical specification. Hierarchical modelling thus reveals the ward-specific variability of respondents' participation in undeclared work. The between-wards variation is a crucial ingredient brought to the fore by a hierarchical specification. It offers an opportunity to examine, for instance, whether respondents belonging to a ward behave similarly as far as participating in undeclared work is concerned, and whether they behave differently compared to respondents from other wards.

In addition, to determine the validity of a hierarchical model one must compare the ratio of between-wards (or between-groups) variation to the overall model variation. If the proportion of between-wards to overall variability is large, using a single-level (i.e. traditional) regression instead of a hierarchical logistic specification will lead to inefficient and inconsistent estimates. Such a situation also implies that the between-wards variability is vital to explaining the overall variability of the phenomenon under investigation (Hox et al., 2017).

Ultimately, it is possible to contextualise this study's model specification in line with regional science, where it has been established that phenomena or objects belonging to a geographical area (a 'ward', in the present case) are expected to behave similarly. It has been suggested that this similarity of behaviour can be attributed mainly to area-specific but unobserved factors that influence the phenomenon under investigation. For instance, a concentration of respondents with similar socioeconomic characteristics in a ward may also explain their similarity in terms of participating in undeclared work. In this study, the area-wide unobserved factors were captured by introducing the random-effects component in the specification, which is elaborated on in detail in the next section. The hierarchical model's estimation procedure is also discussed in the next section.

Model estimation procedure

To facilitate a coherent understanding of the procedure adopted in this study for estimating a hierarchical logistic model, some background is required. In this regard, the study area (Gauteng) was partitioned into P non-overlapping areal units whose geographical boundaries correspond to 509 wards, as they existed in 2018. Respondents in the QoL survey are indexed by $i(i = 1, \dots, N)$, whereas the wards in which these respondents reside are symbolised by $j(j = 1, \dots, P)$. A total of 8 134 respondents who indicated that they participated in either declared or undeclared work were considered in the analysis. It should also be noted that the number of wards is smaller than the number of respondents $(P < N)$, making hierarchical modelling possible.

The dependent variable vector and the matrix of independent variables are associated with respondents, as described above. The dependent variable is a binary indicator, equal to one if a respondent participates in undeclared work $(y_{ij} = 1)$, and zero if they participate in declared work $(y_{ij} = 0)$. Taking the logistic transformation into account means predicting the probability of a respondent falling into a target group – namely, in this case, participating in undeclared work. This probability can be represented as:

$$Prob(y_{ij} = 1) \tag{1}$$

In logistic regression, the logit is used instead of probabilities bound to zero and one. It is important to note that the logit transforms the predicted probability of target-group membership. The main reason for using the logit is to ensure that there is a linear relationship between the independent variables and the likelihood of a respondent participating in undeclared work. Equation (2) represents the logit transformation of a respondent's likelihood of participating in undeclared work. A matrix of

K independent variables is denoted by $X_{ij}^T = (x_{ij1}, \dots, x_{ijK})$.

$$logit(y_{ij} = 1) = ln\left(\frac{Prob(y_{ij}=1)}{Prob(y_{ij}=0)}\right) \tag{2}$$

This paper followed two steps in estimating the hierarchical logistic model, as shown below. At each stage, the model was estimated following the Bayesian context, using the integrated nested Laplace approximation (R-INLA) algorithm programmed in the R-INLA package.[2]

Step 1: Hierarchical null logistic model

Equation (3) represents the null logistic model:

$$\begin{aligned}&logit(y_{ij}=1)=\alpha_{ij}+u_j \\ &u_j \sim N(0,\tau_u) \\ &\pi(\alpha_{ij}) \sim N(\mu_\alpha,\sigma_\alpha^2) \\ &\pi(\tau_u) \sim logGamma(1,10^{-5}),\end{aligned} \tag{3}$$

where σ_{ij} is the overall intercept of log odds of a respondent participating in undeclared work across all wards, and u_j is the intercept specific for ward j. For computation purposes, the algorithm in R-INLA uses the inverse of the variance of the random effect component, $\left(\tau_u=\frac{1}{\sigma_\mu^2}\right)$. This inverse is referred to as 'precision'. However, when the estimates are reported on in the Data section (below), the variance is shown for interpretation, not precision. The term π symbolises the prior distributions. These prior coefficients are included in the estimation of posterior coefficients following the Bayesian approach (for more details on R-INLA priors, see Wang et al., 2018).

Furthermore, Equation (3) has two components, notably the fixed-effects (α_{ij}) and random-effects (u_j) components. Coefficients in the fixed-effects component, in this case the overall model intercept, are interpreted as for any traditional regression. In contrast, the variance of the random-effects element is of great concern in hierarchical modelling, as already indicated. Equation (3) tests the appropriateness of the hierarchical logistic model against a single-level logistic model. Technically, Equation (3) evaluates whether the between-wards variance is significant enough to explain the log-odds of participation in undeclared work. In other words, after estimating Equation (3), the author calculated the intraclass correlation coefficient (ICC), a statistic

[2] The R-INLA package is freely available at R-INLA Project.

that indicates the relative magnitude of the between-wards variance component. It also quantifies the degree of homogeneity within wards, regarding whether a respondent participates in undeclared work or not. Equation (4) is the formula for calculating the ICC from the estimated null hierarchical logistic model:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \left(\frac{\pi^3}{3}\right)}, \tag{4}$$

where $(\pi^3/3) \approx 3.29$ is the standard logistic distribution of the level 1 variance component. A higher value of ICC implies dependence between respondents within wards, whereas a lower ICC shows independence. The conventional threshold to determine whether the between-wards variance is significant or not requires the ICC to be equal to or greater than 0.05 (Heck et al., 2014). This paper concludes that if the estimated ICC is equivalent to or greater than 0.05, then a hierarchical logistic model is necessary, and one must then proceed to the second step below.

Step 2: Hierarchical random intercept logistic model

After confirming the suitability of a hierarchical logistic model over a single-level logistic model, the next step was to estimate what is referred to as a 'hierarchical random intercept' logistic model. Equation (5) represents such a model:

$$\begin{aligned} &logit(y_{ij} = 1) = \alpha_{ij} + X_{ijk}^T \beta + u_j \\ &u_{ij} \sim N(0, \tau_u) \\ &\pi(\alpha_{ij}) \sim N(\mu_\alpha, \sigma_\alpha^2) \\ &\pi(\beta) \sim N(0, 10^6) \\ &\pi(\tau_u) \sim logGamma(1, 10^{-5}) \end{aligned} \tag{5}$$

The meanings of the terms in Equation (5) remain the same as in Equation (3). However, the matrix of independent variables with their associated K slopes (β) vector is now incorporated into the model. Equation (5) predicts the log-odds that *respondent i* in *ward j* participates in undeclared work as a function of the overall intercept (α_{ij}), with selected independent variables that include the variable of interest (i.e. *immigrant*) and the vector of the random effects of wards (u_j). This model allows the intercept to vary between wards. For instance, the intercept of the log-odds that a respondent in a specific ward participates in undeclared work is $\alpha_{ij} + u_j$. The random effects are assumed to be independent, with zero mean and constant variance. A discussion of the data used in this study follows.

Data

Table 1 shows the variables used to estimate the models presented. Some of these variables were constructed based on the fifth wave of the QoL survey (details of the sampling strategies adopted in the QoL survey can be found in Gauteng City-Region Observatory (2019)). These variables were selected as informed by the available data. Most importantly, the selection of control variables is based on the literature. Scholars such as Beręsewicz and Nikulin (2018), as well as Williams and Horodnic (2015a, 2015b, 2015c), have contended that a structuralist exclusion lens reveals the significance of factors such as education, age and gender in understanding the engagement of individuals in undeclared work. Interested readers may consult these works for further elaboration.

Table 1. Description of variables

Variable	Description
Dependent variable	
Undeclared work	A binary variable that indicates whether a respondent participates in undeclared work. It is equal to 1 if they participate in undeclared work; otherwise it is 0.
Independent variables: Individual characteristics	
Immigrant	A binary variable. It is equal to 'Yes' if the respondent was born outside South Africa; otherwise, it is equal to 'No'.
Age	Age of respondent in years at the time of the interview.
Gender	A binary variable. The reference level is 'Male'; otherwise, it is 'Female'.
Ethnicity	A binary variable that indicates the population group to which the respondent belongs. The reference level is 'Black'; otherwise, it is 'White'.
Education	A categorical variable. It indicates the respondent's education level. The reference level is 'Secondary' education; otherwise, it may be either 'No secondary' or 'Post-secondary'.
Independent variables: Household characteristics	

Size	Number of adult equivalents in the respondent's household.
Dwelling	A categorical variable. It indicates the typology of the respondent's dwelling. The reference level is 'Informal'; otherwise, it is 'Formal'.

As stated in the article's introductory section, the dependent variable *undeclared work* was constructed on the basis of the following question in the QoL survey: *What is your employment status?* The questionnaire provided the following six possible answers, and respondents were required to select the one that applied to them: a) Employed full time, formal sector; b) Employed part time, formal sector; c) Employed full time, informal sector; d) Employed part time, informal sector; e) Self-employed, own business, not working from home; and f) Self-employed, own business, working from home. Because the dependent variable is a binary indicator, these responses had to be collapsed into two categories. Respondents who selected c and d as answers were amalgamated into the first category, where the dependent variable takes the value 1 (i.e. *undeclared work* is equal to 1). In contrast, those who chose a and b belonged to the second category, where the dependent variable has the value 0 (i.e. *undeclared work* is equal to 0). Respondents who selected e and f were excluded from the analysis because the focus of the study was on waged work and not on self-employment. Furthermore, the questionnaire used in the QoL made it difficult to determine if self-employed individuals were in the informal or the formal sector. Therefore, self-employment was excluded. However, a follow-up paper will investigate the relationship between self-employment and immigration, as this too is of great concern for Gauteng.

The independent variables were classified into two groups: the first group represented characteristics that apply to respondents as individuals, and the second group showed the characteristics of the respondents' households. In this regard, the independent variable of interest in this paper, *immigrant*, belongs to the individual-level characteristics. It was constructed from the question: *In which province or country were you born?* Respondents were required to indicate the province of their birth, and there was an option to indicate whether or not they were born outside of South Africa. The variable *immigrant* took a value of 1 if a respondent was born in South Africa.

The description of the remainder of the variables reported in Table 2 is straightforward, except for *size*. The question arises whether or not one should consider the number of household members provided by the

QoL survey, irrespective of age. In other words, what constitutes the size of a household? Should one consider only adult members, or any member, regardless of age? This study used 'equivalent persons' instead of 'persons' to record the number of household members. Coincidentally, Question 15_07 of the QoL survey provides information regarding the number of child dependants living in the household with the respondent. For analysis in this study, the actual household size, based on the working definition of 'equivalent person', was determined from information obtained from the QoL survey, as mentioned above.
In effect, the number of children who are household members was first divided by two. After obtaining the corrected number of child members, this number was added to the number of adult members to arrive at the total number of household members to be used as the variable *size*. Even though this number may have shortcomings, the author's view is that it considers the differences in household composition from a size point of view.

Table 2. Summary of categorical variables

Variable	Category	Number	Percentage
Undeclared work	**Yes**	1 997	25
	No	6 137	75
Immigrant	Yes	791	10
	No	7 343	90
Gender	**Male**	4 719	58
	Female	3 415	42
Education	**Secondary**	2 463	30
	No secondary	2 476	30
	Post-secondary	3 195	40
Dwelling	**Informal**	1 392	17
	Formal	6 742	83
Ethnicity	**Black**	7 070	87
	White	1 064	13

Table 2 above provides information on all the categorical variables considered in the paper. The first column of this table ('Variable') shows each variable, while the second ('Category') depicts the categories associated with each variable. The reference categories are displayed in bold font. The proportion of households that fall within each category is shown in the 'Percentage' column. For instance, one can read the first row, which is related to the variable *undeclared work*, as stating that 25% of respondents indicated that they participate in undeclared work. This proportion is sufficient to ensure a certain degree of variability in the distribution of the dependent variable.

According to the survey, around 10% of the respondents are immigrants. From a statistical analysis, this level of variability in the data is acceptable. Furthermore, out of 791 immigrants in Gauteng, Zimbabweans accounted for the highest proportion, at 39%, followed by people from Lesotho at 14%. Among the 791 migrants, it was found that 16% of Zimbabwean immigrants were involved in undeclared work. This means that out of all the Zimbabwean immigrants in Gauteng, 42% were involved in undeclared work. Similarly, among Basotho immigrants in Gauteng, 39% were found to be involved in undeclared work.

Most interesting is the observation that 30% of respondents reported having no education. Mulamba (2020) also reported that a low proportion of household heads in South Africa have secondary education. Since secondary education is considered the basic level of education for one to participate meaningfully in active life, this finding is an indication that a more significant effort is needed to educate all South Africans.

Table 3. Summary statistics of continuous variables

Variable	**Mean**	**Minimum**	**Maximum**	**STD**
Age	38.58	18	88	10.33
Size	2.42	1	13.5	1.34

Table 3 above provides key descriptive statistics regarding the two continuous variables considered in this paper: *age* and *size*. Although these were scaled to avoid the issue of aberrant values in the actual regression analysis, here they are reported without being scaled. Table 3 shows the disparities between respondents that can be attributed to the reported large standard deviations.

Discussion of the Results

Table 4 below presents the Equation (5) estimates, as discussed in the Model Specification section. First, it is noted that the posterior distributions reported in this table include means (second column) and 95% confidence intervals (fourth and sixth columns). Second, Equation (3) was also estimated, and the ICC is reported in the final row of Table 4. As can be seen, the ICC is equal to 0.116, indicating that the hierarchical rather than the single-level model is most suitable for the sample. This ICC is greater than the set threshold of 0.05, indicating that the 12% variance in the log-odds of a respondent participating in undeclared work can be attributed to differences between wards.

Table 4. Posterior estimates of the hierarchical random intercept logistic model

Variable	Mean	SD	25%	50%	97.5%
Constant	**-0.952**	0.083	-1.116	-0.951	-0.789
Immigrant: Yes	**0.598**	0.180	0.242	0.599	0.947
Immigrant: Yes and Female	0.225	0.186	-0.138	0.225	0.59
Immigrant: Yes and No secondary	-0.037	0.203	-0.434	-0.038	0.363
Immigrant: Yes and Post-secondary	-0.120	0.286	-0.692	-0.116	0.433
Age	**-0.100**	0.029	-0.157	-0.100	-0.042
Gender: Female	**0.207**	0.062	0.086	0.207	0.328
Ethnicity: White	**-0.807**	0.137	-1.081	-0.805	-0.544
Education: No secondary	**0.756**	0.072	0.614	0.756	0.898
Education: Post-secondary	**-1.022**	0.084	-1.189	-1.022	-0.857
Size	0.028	0.028	-0.027	0.028	0.084
Dwelling: Formal	**-0.344**	0.073	-0.487	-0.344	-0.201

SD of random effects	0.485	0.043	0.401	0.484	0.571
ICC	0.116				
Number of respondents (Level 1)	8 134				
Number of wards (Level 2)	529				

Posterior mean estimates are expressed in logit. Estimates in bold mean they are not statistically significant.

With the hierarchical model confirmed as suitable for the sample, as discussed in the Model Specification section, the next step was to estimate Equation (4). For ease of flow, the discussion on the independent variable of interest (*immigrant*) is presented first. As illustrated in Table 4, the posterior mean of *Immigrant: Yes* and its associated 95% confidence interval are 0.598 and (0.242, 0.947) respectively. First, this finding indicates the presence of a positive relationship between being an immigrant and participating in undeclared work in Gauteng, thus confirming the structuralist exclusion hypothesis (as a reminder, the structuralist view holds that immigrants or other marginalised individuals are prone to participating in undeclared work). This result also implies that some societal factors prompt more immigrants than non-immigrants to participate in undeclared work. Based on this finding, policymakers are encouraged to identify these factors to ensure immigrant workers participate in declared work that benefits the economy. Informed policy formulation is also essential to address the poor working conditions to which workers – particularly immigrants, who also face other challenges – are often exposed when participating in undeclared work. One way the government can encourage formalisation is by simplifying the business registration process. In addition, the government could ensure that the labour rights of immigrants are protected, regardless of whether they work in the formal or the informal sector. The government should prioritise easing the process of qualification transferability to facilitate the recruitment of skilled and documented immigrants for formal employment opportunities.

Second, the likelihood of an immigrant respondent in Gauteng – regardless of the ward in which they reside – participating in undeclared work is 1.81 (exp (0.598)) times higher than for a non-immigrant respondent, on average and all other things being equal. This result reinforces the structuralist exclusion hypothesis, as discussed above. Relative to immigrants, therefore, non-immigrant respondents participate less in undeclared work because they naturally experience fewer exclusions in the labour market.

Two interactive variables were created and introduced in Equation (5) for estimation purposes, as shown in Table 4 above. The first interaction is between *immigrant* and *gender*, while the second is between *immigrant* and *education*. These interactions were considered mainly to assess whether the likelihood of an immigrant participating in undeclared work was significantly different depending on gender and education level. As shown in Table 4 above, the posterior estimates of the relevant interaction variables were not statistically significant. This finding demonstrates that immigrants are predisposed to participating in undeclared work in Gauteng, regardless of gender, education level or ward of residence.

The posterior mean estimates of the control variables were all statistically significant, except for those related to *size*. One implication of these findings is that a respondent's household size is not a determinant of participation in undeclared work in Gauteng province. Furthermore, the odds relating to a respondent's age range between -0.157 and -0.04 of the 95% confidence interval. This finding is an indication that the probability that a respondent would participate in undeclared work decreases (increases) as their age increases (decreases), all things being equal. In other words, older people in the province are less predisposed to participating in undeclared work than younger people. Moreover, female respondents are more likely to participate in undeclared work than their male counterparts. This finding again reinforces the hypothesis that women are excluded from declared work.

The reported posterior estimates related to the variable *Ethnicity: White* reveal another dimension of disparity in Gauteng. For instance, the posterior mean estimate shows that the odds of a respondent belonging to the white population group participating in undeclared work are 0.446 times lower than for a respondent from a non-white group. Consequently, white people are better off than non-white people in terms of participating in declared work in the province, all things being equal. This finding is not surprising given South Africa's historical background, in that in many respects the apartheid government privileged white people to the detriment of black people. White people, for instance, had access to better education, which is a determinant of participation in the formal labour market. This is discussed below.

Most importantly, it is noted that the posterior mean estimates related to the variables *Education: No secondary* and *Education: Post-secondary* are positive and negative respectively. For the former, this finding intuitively implies that participation in undeclared work is positively associated with a lack of secondary education. In contrast, respondents with post-secondary education are less likely to participate in undeclared work than those with only secondary education, all other things being equal. Another exclusion aspect is confirmed by the finding relating to respondents' dwelling typologies. As shown in Table 4 above, respondents residing in formal dwellings were less likely to participate in undeclared work than those residing in informal dwellings.

Thus far, the discussion in the preceding paragraphs of the interpretation of posterior estimates has focused on the fixed-effects component. For instance, understanding the fixed effects related to the variable *immigrant* helps to disentangle the behaviour of an immigrant respondent who participates (or not) in undeclared work, regardless of his or her ward of residence. On the contrary, if one seeks to understand the behaviour of an immigrant respondent residing in *ward j* (regarding their participation in undeclared work), the posterior mean coefficient would be $0.598 + u_j$. This figure corresponds to the mean of the fixed-effect component (of *immigrant*), augmented by the estimated random effect specific to *ward j*.

This study's empirical findings (based on the information obtained from the QoL survey) consistently demonstrate the structuralist hypothesis of exclusion for Gauteng. Together with the findings for other predictors, this result could be an essential ingredient in informing the formulation of labour and immigration policies, since a rigorous methodological process was followed to estimate the relationship between immigrants and their participation in undeclared work.

Conclusion

This paper empirically tested and determined the degree to which immigrants participate in undeclared work in Gauteng, South Africa. This topic is particularly relevant to Gauteng, since the province attracts many immigrants from neighbouring countries seeking better opportunities. The empirical analysis tested the structuralist hypothesis of exclusion using data from the fifth wave of the QoL survey. This paper adopted hierarchical logistic regression as a suitable modelling approach to achieving its objective. Consequently, it considered group effects when analysing the economic behaviour of the respondents in different geographical settings.

A rigorous diagnostic process determined that the hierarchical random intercept model was appropriate for the sample data. Among other conclusions, this finding implies that respondents from the same ward exhibit similar behaviour as far as participating in undeclared work is concerned. The empirical results also confirm the validity of the structuralist exclusion hypothesis. It was established that an immigrant is more likely than a non-immigrant to participate in undeclared work in the province.

References

Akintola, O. O., & Akintola, O. (2015). West Africans in the informal economy of South Africa: The case of low skilled Nigerian migrants. *Review of Black Political Economy*, *42*(4), 379–398. https://doi.org/10.1007/s12114-015-9211-8

Beręsewicz, M., & Nikulin, D. (2018). Informal employment in Poland: An empirical spatial analysis. *Spatial Economic Analysis*, *13*(3), 338–355. https://doi.org/10.1080/17421772.2018.1438648

Blaauw, P. F., & Pretorius, M. (2022). "I am 30 and I have nothing": The context of reception and the lived experiences of foreign migrants working as car guards in Johannesburg's West Rand. *GeoJournal*, *88*, 121–135. https://doi.org/10.1007/s10708-022-10594-8

Charman, A., & Petersen, L. (2015). A transnational space of business: The informal economy of Ivory Park, Johannesburg. In J. Crush, A. Chikanda, & C. Skinner (Eds.), *Mean streets: Migration, xenophobia and informality in South Africa* (pp. 78–99). Southern African Migration Programme.

Crush, J., Skinner, C., & Chikanda, A. (2015). *Informal migrant entrepreneurship and inclusive growth in South Africa, Zimbabwe and Mozambique*. Southern African Migration Programme Policy Series No. 68.

Foster, J., Chasomeris, M., & Blaauw, D. (2021). Re-visiting car guarding as a livelihood in the informal sector. *Development Southern Africa*, *38*(5), 769–784. https://doi.org/10.1080/0376835X.2021.1891862

Gashi, A., & Williams, C. C. (2019). Evaluating the prevalence and distribution of unregistered employment in Kosovo: Lessons from a 2017 survey. *South East European Journal of Economics and Business*, *14*(1), 7–20. https://doi.org/10.2478/jeb-2019-0001

Gauteng City-Region Observatory. (2018). *Quality of Life Survey 2017-2018, Round 5. Version 1.1*. https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/766

Gauteng City-Region Observatory. (2019). *Field report for Quality of Life Survey V*. https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/766/related-materials

Heck, R. H., Thomas, S., & Tabata, L. (2014). *Multilevel modeling of categorical outcomes using IBM SPSS*. Routledge.

Hox, J. J., Moerbeek, M., & Van de Scoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd edition). Routledge.

Kalitanyi, V. (2010). African immigrants in South Africa: Job takers or job creators. *SAJEMS - South African Journal of Economic & Management Sciences - Suid-Afrikaanse Tydskrif vir Ekonomiese en Bestuurswetenskappe*, *13*(4), 376–390.

Lekabe, T. (2021, September 22). Mashaba vows to tackle service delivery, illegal immigration, corruption if elected. *The Citizen*. https://www.citizen.co.za/news/south-africa/elections/local-2021/mashabas-actionsa-promises-to-tackle-service-delivery-corruption-illegal-immigration-22-september-2021/

Mafata, M. (2022, February 2). Immigrant traders in Gauteng fear for their safety after wave of xenophobic attacks. *GroundUp*. https://www.news24.com/news24/southafrica/news/immigrant-traders-in-gauteng-fear-for-their-safety-after-threats-of-xenophobic-attacks-20220202

McEwen, H., & Leiman, A. (2008). The car guards of Cape Town: A public good analysis. *Minutes of the Proceedings*, *81*(1885), 25.

Mulamba, K. C. (2020). Relationship between education and households' electricity-saving behaviour in South Africa: A multilevel logistic analysis. *Economics and Policy of Energy and the Environment*, *2*, 51–74. https://doi.org/10.3280/efe2020-002003

OECD/ILO. (2018). *How immigrants contribute to South Africa's economy*. https://doi.org/10.1787/9789264085398-en

Pretorius, M. & Blaauw, D. (2019). Kwesbare groepe in die informele ekonomie : 'n Gevallestudie van motorwagte in Johannesburg se Wesrand. *Tydskrif Vir Geesteswetenskappe, 59*(4), 642–656.

Schneider, F. (2008). The economics of the hidden economy. Edward Elgar Publishin.

Steyn, F. (2018). Om na uitbuiting toe te vlug: Die geval van immigrante wat as motorwagte werk (*Fleeing to exploitation: The case of immigrants who work as car guards*). *Tydskrif vir Geesteswetenskappe*, *58*(4–2), 924–939.

Theodore, N., Blaauw, D., Pretorius, A., & Schenck, C. (2017). The socioeconomic incorporation of immigrant and native-born day labourers in Tshwane, South Africa. *International Migration*, *55*(1), 142–156. https://doi.org/10.1111/imig.12311

Thompson, D. K. (2016). Risky business and geographies of refugee capitalism in the Somali migrant economy of Gauteng, South Africa. *Journal of Ethnic and Migration Studies*, *42*(1), 120–135. https://doi.org/10.1080/1369183X.2015.1073580

Valencia, O., Angarita, M., Santaella, J., & Castro, M. (2020). *Do immigrants bring fiscal dividends? The case of Venezuelan immigration in Colombia* (IDB Working Paper Series No. IDB-WP-1170). https://doi.org/10.18235/0002993

Wang, X., Yue, Y. R., & Faraway, J. J. (2018). *Bayesian regression modeling with INLA*. Taylor & Francis.

Williams, C. C., & Efendic, A. (2021). Evaluating the relationship between marginalization and participation in undeclared work: Lessons from Bosnia and Herzegovina. *Journal of Southeast European and Black Sea, 21*(3), 481–499. https://doi.org/10.1080/14683857.2021.1928419

Williams, C. C., & Horodnic, I. (2015a). Are marginalised populations more likely to engage in undeclared work in the Nordic countries? *Sociological Research Online, 20*(3), 15–29. https://doi.org/10.5153/sro.3719

Williams, C. C., & Horodnic, I. (2015b). Marginalisation and participation in the informal economy in Central and Eastern European nations. *Post-Communist Economies*, *27*(2), 153–169. https://doi.org/10.1080/14631377.2015.1026686

Williams, C. C., & Horodnic, I. (2015c). Marginalisation and participation in the informal economy in Central and Eastern European nations. *Post-Communist Economies*, *27*(2), 153–169. https://doi.org/10.1080/14631377.2015.1026686

Williams, C. C., & Horodnic, I. A. (2015d). Self-employment, the informal economy and the marginalisation thesis: Some evidence from the European Union. *International Journal of Entrepreneurial Behaviour and Research*, *21*(2), 224–242. https://doi.org/10.1108/IJEBR-10-2014-0184

Williams, C. C., & Round, J. (2010). Explaining participation in undeclared work: A result of exit or exclusion? *European Societies*, *12*(3), 391–418. https://doi.org/10.1080/14616691003716910

World Bank Group. (2018). *Mixed migration, forced displacement and job outcomes in South Africa*. World Bank. https://doi.org/10.1596/30158

Annexure

Non-Technical Note on the Hierarchical Logistic Model

This note aims to explain a hierarchical logistic model in simple terms for readers who may not have a strong statistical background. To begin with, a logistic model is a regression model used when the outcome or response (dependent) variable is categorical. This means that the variable takes values such as 'Yes' and 'No', 'Male' and 'Female', 'Old' and 'Young', and 'Failure' and 'Success'. The model helps predict the probability of something happening, such as success or failure, or a yes or no answer. In the case of this study, the aim was to predict the probability of an immigrant participating in undeclared work in Gauteng.

Let us shift the focus to the term 'hierarchical', which pertains to the data structure being scrutinised. The data is arranged in a nested or hierarchical structure. For instance, students may be nested within classrooms, which are then nested within schools. In the case of this study, individuals who were interviewed were nested within wards. This structure indicates that observations (individuals) within the same group (wards) may exhibit more similarities between each other than when observed in relation to other groups. Therefore, when modelling, one must consider this hierarchical structure in the data.

The integration of the logistic model and hierarchical concepts results in a powerful tool called the 'hierarchical logistic model'. This model combines logistic regression with hierarchical data structure analysis, allowing for the examination of categorical outcomes (such as the probability of engaging in undeclared work), while accounting for the nested nature of the data. By analysing how individual and group-level factors influence the outcome, this model provides valuable insights into how both individual and group characteristics influence the likelihood of a given event, recognising that individuals are part of larger groups or wards. Essentially, a hierarchical logistic model is an effective way to understand the influence of individual and group-level (ward-level) factors on participation in undeclared work in Gauteng.