IHS Working Paper 34
August 2021

On the willingness to exit street hawking

Adusei Jumah
Ernest Somua-Wiafe
Barnabas Apom

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

Author(s)

Adusei Jumah, Ernest Somua-Wiafe, Barnabas Apom

Editor(s)

Robert M. Kunst

Title

On the willingness to exit street hawking

Institut für Höhere Studien - Institute for Advanced Studies (IHS)
Josefstädter Straße 39, A-1080 Wien
T +43 1 59991-0
F +43 1 59991-555
www.ihs.ac.at
ZVR: 066207973

On the willingness to exit street hawking

Adusei Jumah[a*], Ernest Somua-Wiafe[a], Barnabas Apom[b]

[a]Central University, Accra, Ghana

[b]Researchlime Ltd, Accra, Ghana

[*] Corresponding author.
Email-mail address: ajumah@central.edu.gh (A.Jumah)

On the willingness to exit street hawking

Abstract:

This paper uses a generalized estimating equations approach to investigate the willingness of street hawkers in Accra to exit their trade when offered viable alternatives. We observe a direct relationship between willingness to exit hawking and age. Willingness to exit hawking declines over increasing experience. Hawkers who have encountered arrest are less likely to exit. Also, hawkers who received financial assistance to facilitate their business are less likely to exit. Plus, monthly savings from hawking may indirectly influence a rejection of the offer to start an alternative trade. Finally, the willingness to quit hawking is not significantly influenced by gender, sales revenue, and the road description in the model. The results of the analysis could help improve the success of future intervention programmes.

Keywords: Street hawking, GEE, logit, willingness to exit

JEL Classification. C35, E26

Introduction

In most developing countries, the urban poor survives by working in the informal sector. Poverty and lack of gainful employment in the rural areas and the smaller towns drive large numbers of people to the cities for work and livelihood (Pappeswari and Rajalakshmi, 2014; Panwar and Garg, 2015). These people generally possess soft skills and lack the education required for the organized sector's better-paid jobs. Moreover, permanent protected jobs in the organized sector are shrinking. This is partly due to the increasing emphasis on governments' reduced role in directly creating employment (Bryceson and Potts, 2006); hence, even those with requisite skills cannot find proper jobs (Bhowmik, 2000). For the urban poor, hawking is one way to earn a livelihood, as it requires little financial input, and the skills involved are low (Pappeswari and Rajalakshmi, 2014).

While developing countries have a long-standing history of street hawking, globalization and trade liberalization in recent decades have made available various kinds of products on the streets of these countries. The rise in cheap Chinese imports has reinforced this trend. And together with the expanding number of motor vehicles without the corresponding increase in access roads or widening of streets, these dynamics have intensified traffic congestions, with the most congested points serving as ready markets for street hawkers (Steel, Ujoranyi and Owusu, 2014).

For the most part, developing countries seek to regulate street hawkers. However, they lack the means to do so (Matthews and Vega, 2012). Thus, even though institutional regulations exist (Matthew and Vega, 2012), they tend not to be operational. The virtual absence of clear legal frameworks to accommodate street vendors in urban planning exposes them to workplace risks (insecurity; harassment; merchandise confiscations as well as evictions) (te Lintelo, 2010). In the context of that challenge, street vendors in many cities operate in uncertain work environments (Roever and Skinner, 2016).

Street hawking is tedious and dangerous business (Anjaria, 2006). It involves long hours of work. By contributing to vehicular and pedestrian congestion, street vendors may cause traffic accidents, increase vehicle-generated air pollution levels, and impede the flow of police, fire, ambulance, and other emergency vehicles. Crowded sidewalks, vendors in the roadway, and pedestrians displaced onto the roadway may block motorists' sightlines at intersections. The lively activity of street sales may distract motorists from their driving (Bromley, 2000).

When a business activity, despite its riskiness, is so widespread, there must be strong forces behind its persistence. An appreciation of these forces is useful in determining which particular impacts of policy interventions are especially important (see also Parikh et al., 1988). The ease of market entry, reliance on indigenous resources; small-scale operation; labour intensive (low capital requirement per worker); unsophisticated technologies; unregulated and competitive market (Dey and Dasgupta, 2010) are the main reasons for street hawking persistence. Hawkers often cite the lack of other employment opportunities as the reason for doing this rather unpleasant work (Anjaria, 2006). If other, more viable alternatives to street hawking can be found, limiting the menace introduced by street hawking should be comparatively easy. Yet street hawking persists.

Many governments have tried to pull selected street vendors into programs to promote entrepreneurship through business education, low-interest credit, and public health training. Such support programs usually target vendors who sell primarily to tourists and middle- to upper-income groups. However, in most cases, participation rates are low, and both extension workers and street vendors have numerous complaints (Bromley, 2000).

There is no real theory or code of ethics associated with regulating or promoting street vendors (Bromley, 2000). Studies on street hawking are primarily descriptive analysis of survey data. Championed by Bhowmik (2003; 2005; 2010a; 2010b; 2012 among others) these studies are mostly concentrated on Asia, particularly on India. Increasingly, several studies are focusing on African countries. For example, Women in Employment, Globalising and Organising (WIEGO) — an international network — has worked closely with StreetNet in leveraging funds for research and policy dialogue on street trade in six African countries, namely South Africa, Ghana, Cote d'Ivoire, Uganda, Zimbabwe and Kenya (Mitullah, 2004). All these studies, however, cluster around the six common problems of street traders around the world categorized by the 1995 Bellagio International Declaration of Street Vendors, which are: lack of legal status and right to vend, lack of space or poor location, restriction on licensing, cost of regulation, harassment, bribes, confiscation and evictions, lack of services and infrastructure and lack of representation or voice. Virtually, no rigorous empirical research exists on street hawking. This paper seeks to fill that gap.

We investigate the willingness of street hawkers in Accra to exit their business activities when offered viable alternatives by means of generalized estimating equations (GEE) modelling. Pekár and Brabec (2018) discussed desirable properties of the marginalized GEE models as recommended for behavioural related responses with distributions typically generated from the family of exponential distributions. Also, for designs that are sampled on repeated, nested or longitudinal outcomes, the GEE approach is regarded as an alternative to the random effect approach in modeling such correlated responses over population averaged effects.

Ballinger (2004) also demonstrated the use of the GEE approach on normal responses for nested subjects and for longitudinal count responses over time-variant and invariant covariates taking into account the correlations that existed within the responses. Aloisio et al. (2014) utilized the GEE approach in analysing partially correlated outcomes with missing data using the multiple imputation mechanism. Investigating the prevalence of eating disorder symptoms in adolescents, they observed fairly similar estimates for both missing-at random and the missing completely at random models under the GEE specification and considered the missing-at random imputation as superior. Likewise, Huh, Flaherty, and Simoni (2012) in their research identified the logistic GEE with robust standard errors as preferable to the classical analysis of variance (ANOVA) test in investigating the effects of adherence to HIV interventions.

According to the 1995 Bellagio International Declaration of Street Vendors, widespread street trading is largely the result of uneven distributions of wealth and although the declaration is more than twenty years old, it still has plenty of sound advice for African cities (Agbo, 2017). Accra, like many major cities in Africa, faces increasing traffic congestion, pedestrians, and traders around its central business district. Repeated attempts to evict street traders have proven unsustainable (Steel, Ujoranyi and Owusu, 2014). Presumably, former government benevolent interventions might have failed or suffered low participation rates because they had not considered hawkers' willingness to exit their businesses, given the persistence of street hawking. Measuring street hawkers' willingness to exit hawking will give policy makers flexibility to target multiple segments of street vendors at different incentive levels. The results of the analysis could help improve the success of future intervention programmes.

The organization of the paper is as follows: Section 2 expounds the survey design and model employed in the study; Section 3 interprets the empirical results and Section 4 concludes.

2. Methodology

2.1 The generalized estimating equations (GEE)

The generalized estimating equations (GEE) specification is a special generalized linear model (GLM) tailored to assess marginal exposure effects while accounting for the covariance structure within responses (Liang and Zeger, 1986). As a result, GEE generates population-averaged (PA) estimates which makes the generalization of effects plausible. Both the logit model and the probit model have been frequently used in obtaining marginal probability estimates over exposure effects given dichotomous responses and hence with similar link functions, we discuss the model framework for our outcome variable.

Suppose y_{ij} denotes a binary outcome which takes on discrete values, 0 and 1 with $i = 1, 2, \dots, m$ identifying subjects and $j = 1, 2, 3, \dots, n_i$ the observations reiterated at different time periods. Then for $\boldsymbol{Y_i} = \left(y_{i1}, y_{i2}, \dots, y_{in_i}\right)$ the set of correlated binary responses in the i^{th} subject with a corresponding vector of p covariates $\boldsymbol{X_i} = \left(x_{ij1}, x_{ij2}, \dots, x_{ijp}\right)$. The GEE model framework for the logit probability function is given as

$$\boldsymbol{\pi}_i{}^{\text{PA}} = P\left(y_{ij} = 1 | \boldsymbol{X_i}\right) = \text{expit}\left(\beta_0{}^{\text{PA}} + \boldsymbol{X_i}\beta_i{}^{\text{PA}}\right)$$

$$y_{ij} \sim \text{Bernoulli}$$

where $\boldsymbol{\pi}_i{}^{\text{PA}}$ denotes the expectation of $\boldsymbol{Y_i}$, and $\beta_i{}^{\text{PA}}$ indicates a vector of coefficients for $\boldsymbol{X_i}$. As well, suppose that $\boldsymbol{V}_i = \text{diag}[\boldsymbol{\pi}_i{}^{\text{PA}} \times (1 - \boldsymbol{\pi}_i{}^{\text{PA}})]$ signifies a diagonal matrix of variances and the correlation matrix describing the within-subject correlation structure for the i^{th} subject is represented by $\boldsymbol{R}_i = (\boldsymbol{r}_{st})_{n_i \times n_i}$ with its elements $\boldsymbol{r}_{st} = \begin{cases} 1 & \text{if } s = t \\ \rho_{st} & \text{otherwise} \end{cases}$, then we can establish the subject-level variance-covariance matrix as $(\boldsymbol{C}_i)_{n_i \times n_i} = (\boldsymbol{V}_i)^{0.5} \boldsymbol{R}_i (\boldsymbol{V}_i)^{0.5}$ where $(\boldsymbol{V}_i)^{0.5} \times (\boldsymbol{V}_i)^{0.5} = \boldsymbol{V}_i$.

For m hawkers with each subject generating n_i responses and $p+1$ regression coefficients $\beta_k{}^{PA}$; $k = 0,1,2,\dots,p$, then the $p+1$ score-like estimating equations with solutions $\hat{\beta}_k{}^{PA}$ are obtained by solving the relation,

$$\sum_{i=1}^{m} \boldsymbol{D}_i{}^T \, [\boldsymbol{C}_i]^{-1} [\boldsymbol{Y}_i - \boldsymbol{\pi}_i{}^{\text{PA}}] = 0$$

where $\boldsymbol{D}_i{}^T = \boldsymbol{X}_i{}' \boldsymbol{V}_i$ with $\boldsymbol{X}_i$ as an $n_i \times (p+1)$ design matrix, and $[\boldsymbol{Y}_i - \boldsymbol{\pi}_i{}^{\text{PA}}]$ specifying the residual terms. The correlation terms and regression coefficients are estimated concurrently in an iterative manner using the iteratively reweighted least squares (IRLS) algorithm up until convergence is realized.

Consequently, the choice of correlation structure could play a significant role in the efficiency of the model since the estimation of standard errors and predictor effects include a specified covariance structure. To ward off over-parameterization, the model imposes a common correlation structure for all within-subject outcomes. An unstructured correlation structure with distinct off-diagonal correlation parameters for within-subject correlations, preserves the order of response without considering the differences in time. An autoregressive (AR) correlation structure also assumes constant time intervals and accounts for missing within-cluster outcomes by making an allowance for the time intervals. The correlation between any two responses from the same individual at different occasions j_1 and j_2 within $\boldsymbol{R}_i$ is of the form $\rho_{j_1 j_2} = \rho^{|j_1 - j_2|}$. Furthermore, in specifying within-subject correlation structure, the form may be simplified by constraining within-subject correlations to a shared parameter $\rho_{j_1 j_2} = \rho$ at all levels.

In measuring variability of GEE coefficients, two general techniques are considered: a model-based approach and an empirical approach. The empirical approach is more robust as compared with the model-based approach since it utilizes both the 'working' correlation and the observed correlation components whereas the former only makes use of the 'working' correlation component. Thus, a misspecification of the correlation structure may well affect the model-based estimates negatively whilst such an error may have no significant effect on the empirical variance estimates. So, by comparison, the robust estimator is much desirable as compared with the model-based estimator. The empirical covariance is obtained as

$$\mathrm{Cov}\left(\hat{\beta}_k{}^{PA}\right) = \left(\sum_{\mathrm{i=1}}^{\mathrm{m}} \widehat{\boldsymbol{B}}_i\right)^{-1} \left(\sum_{\mathrm{i=1}}^{\mathrm{m}} \widehat{\boldsymbol{M}}_i\right) \left(\sum_{\mathrm{i=1}}^{\mathrm{m}} \widehat{\boldsymbol{B}}_i\right)^{-1}$$

where $\widehat{\boldsymbol{B}}_i = \boldsymbol{D}_i{}^T \boldsymbol{C}_i{}^{-1} \boldsymbol{D}_i = \boldsymbol{X}_i{}' \boldsymbol{V}_i \left(\boldsymbol{V}_i{}^{0.5} \boldsymbol{R}_i \boldsymbol{V}_i{}^{0.5}\right)^{-1} \boldsymbol{V}_i \boldsymbol{X}_i{}'$ and the term $\widehat{\boldsymbol{M}}_i$ is also formulated as $\widehat{\boldsymbol{M}}_i = \boldsymbol{D}_i{}^T \boldsymbol{C}_i{}^{-1} \boldsymbol{A}_i\, \boldsymbol{C}_i{}^{-1} \boldsymbol{D}_i = \boldsymbol{X}_i{}' \boldsymbol{V}_i \left(\boldsymbol{V}_i{}^{0.5} \boldsymbol{R}_i \boldsymbol{V}_i{}^{0.5}\right)^{-1} \boldsymbol{A}_i \left(\boldsymbol{V}_i{}^{0.5} \boldsymbol{R}_i \boldsymbol{V}_i{}^{0.5}\right)^{-1} \boldsymbol{V}_i \boldsymbol{X}_i{}'$ with $\boldsymbol{A}_i = (a_{st})_{n_i \times n_i}$ such that $a_{st} = (\boldsymbol{Y}_i - \boldsymbol{\pi}_i{}^{\mathrm{PA}}) \times (\boldsymbol{Y}_i - \boldsymbol{\pi}_i{}^{\mathrm{PA}})^T$ is a product of the empirical residuals.

As part of its properties, GEE estimates are identified as both consistent and asymptotically normal with mean $\beta_k{}^{\mathrm{PA}}$ and variance-covariance of $\mathrm{Cov}\left(\hat{\beta}_k{}^{PA}\right)$ for large samples. Based on the asymptotic properties of the parameter estimates and the estimation techniques employed, the Wald test can be utilized for the global test of significance and testing for the significance of individual covariates within the model.

The odds ratio (OR) estimated at $\widehat{\mathrm{OR}} = \exp\left(\hat{\beta}_i^{\ PA}\right)$ with a confidence interval $(1-\alpha)100\%\ \mathrm{CI} = \exp\left\{\hat{\beta}_i^{\ PA} \pm z_{1-0.5\alpha} \times \mathrm{s.e}\left(\hat{\beta}_k^{\ PA}\right)\right\}$ which is based on asymptotic properties of the sampling distribution of OR are used for inferences.

Another alternative to the logit link function in the model specification is to use the inverse Gaussian function with an initial variance of π^3 and so for good measure, we specify a GEE probit model and compare its fit with the GEE logit model.

Most of the challenges of GEE models are with aspects of model diagnostics and so to address the issue, analysts usually sort to goodness of fit inferences from uncorrelated modelling approaches as justification. We apply an extension of the Hosmer Jr. and Lemeshow (1980) test proposed by Horton et al. (1999) for goodness of fit assessments for the optimal GEE model. Additionally, we explore extensions of well-established Efron (1978), Cox-Snell (1968), McFadden (1974), Cragg and Uhler (1970), and Ben-Akiva and Lerman (1985) pseudo R-squared estimators that functionally work with the maximum quasi-likelihood for GEE models as reported by Deroche (2015).

Analogous to the Akaike Information Criterion (AIC) model comparison, the Quasi-likelihood Information Criteria, $QIC_p(\boldsymbol{R})$ and $QIC_u(\boldsymbol{R})$ proposed by Pan (2001) may be utilized in selecting the best GEE model amongst similar choice models with different specifications of within-subject correlation structure and covariate terms. Unlike the $QIC_p(\boldsymbol{R})$ measure, $QIC_u(\boldsymbol{R})$ approximately penalizes for the number of model parameters with no emphasis on the choice of correlation structure therefore such a measure is useful in selecting the best subset of covariates. An alternative $QIC_{HH}(\boldsymbol{R})$ measure that alters the covariance matrix of the independence model in obtaining the penalty term under the independent estimates could be also useful in selecting the best correlation structure amongst other options (Hardin and Hilbe, 2013). Usually, the model which is smallest $QIC_p(\boldsymbol{R})$, $QIC_u(\boldsymbol{R})$ and $QIC_{HH}(\boldsymbol{R})$ is preferable.

As defined by McCullagh and Nelder (1989), the quasi-likelihood function is given as

$$Q(y;\pi,\phi) = \int_{y}^{\pi} \frac{y-\pi^*}{\phi f(\pi^*)} d\pi^*$$

where $E(y) = \pi$, and $Var(y) = \phi f(\pi)$ with $f(\pi)$ representing a function of the expected value, and ϕ denotes a dispersion scalar factor which depends on the sample size and the number of covariates. Thus, we can express the QIC measures as:

$$QIC_p(\boldsymbol{R}) = -2Q\left(y;\hat{\beta}_{\boldsymbol{R}},\hat{\phi}\right) + 2\mathrm{tr}\left(\widehat{\Omega}_I\left(\hat{\beta}_{\boldsymbol{R}}\right)\hat{C}_{\boldsymbol{R}}\right)$$

$$QIC_u(\boldsymbol{R}) = -2Q\left(y;\hat{\beta}_{\boldsymbol{R}},\hat{\phi}\right) + 2K$$

$$QIC_{HH}(\boldsymbol{R}) = -2Q\left(y;\hat{\beta}_{\boldsymbol{R}},\hat{\phi}\right) + 2\mathrm{tr}\left(\widehat{\Omega}_I\left(\hat{\beta}_{\boldsymbol{I}}\right)\hat{C}_{\boldsymbol{R}}\right)$$

with ϕ estimated at

$$\hat{\phi} = \frac{1}{\left(\sum_i^m n_i\right) - p} \sum_i^m \sum_j^{n_i} \left(y_{ij} - \hat{\pi}_{ij}{}^{\mathrm{PA}}\right) \times \left(y_{ij} - \hat{\pi}_{ij}{}^{\mathrm{PA}}\right)$$

where $Q\left(y;\hat{\beta}_{\boldsymbol{R}},\hat{\phi}\right)$ is the quasi-likelihood with independent within-subject responses, I is an identity matrix, K denotes the number of parameters within the formulated model, $\hat{C}_{\boldsymbol{R}}$ represents the empirical covariance matrix with robust estimates, $\widehat{\Omega}_I\left(\hat{\beta}_{\boldsymbol{R}}\right)$ signifies the inverse of the model-based covariance matrix under the independence assumption evaluated at $\hat{\beta}_{\boldsymbol{R}}$ and $\widehat{\Omega}_I(\hat{\beta}_{\boldsymbol{I}})$ represents the inverse of the model-based covariance matrix under the independence assumption evaluated at $\hat{\beta}_{\boldsymbol{I}}$. Thus, $QIC_p(\boldsymbol{R})$ and $QIC_{HH}(\boldsymbol{R})$ would only be equivalent when $\widehat{\Omega}_I(\hat{\beta}_{\boldsymbol{I}}) = \widehat{\Omega}_I(\hat{\beta}_{\boldsymbol{R}})$.

Even though the statistical diagnostics and comparison of model estimates in some cases may seem to favour the 'naive' model, the precedence is given to one of the models which captures the correlation structure fundamentally associated with the repeated responses in the survey design.

2.2 Research Design

2.2.1 Survey

Structured questionnaires were administered to street hawkers in and around major locations across the Accra Metropolitan Area (AMA). These hawking spots comprised mainly of bus stops, toll booths, highway stops and stations where vendors mostly operate on a daily basis. At each of the sites, hawkers were randomly identified, and enumerators were assigned to each respondent for interviewing. The data collection phases and timelines were as follows:

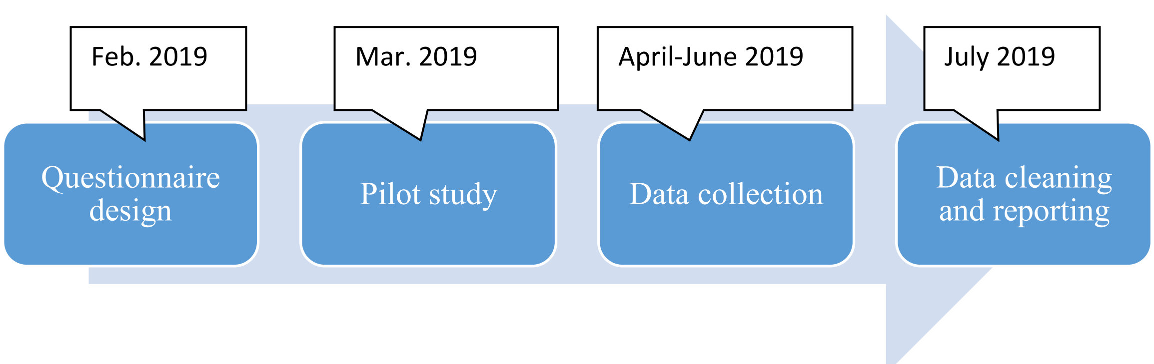

Figure 1: *Data Collection Phases and Timelines*

2.2.2 Sampling and Data Collection

Sampling and selection of respondents were purely random on site. Out of those who were selected, close to 67% elected to participate in the survey. Respondents were mostly either in motion or occasionally at rest. The questionnaire was administered at the moments where traffic flow was not at a standstill and hawkers were retreating to a vantage point to maximize chances of making more sales while competing for space with other hawkers when traffic comes to a halt.

Data collection was carried out via Computer Aided Personal Interviewing (CAPI) that uses mobile devices deployed on an Android-based Open Data Kit (Inyst). Inyst is a mobile data collection tool that enables enumerators to capture responses to questions from respondents and submits surveys to a cloud server as soon as they are completed. Thus, Inyst is a key control tool that helps in ensuring efficiency and adherence to the sampling etiquettes while enabling supervisors to track interviews from the enumerators.

Enumerators visited each selected spot, obtained the consent of hawkers to participate in the survey, and ensured that the information they provided was kept confidential. The enumerators administered the questionnaires, recorded accurate responses, and coded them accordingly, ensuring that the survey was complete and performed accuracy checks whenever applicable. The enumerators then transmitted the responses to a central server over the internet or wireless network.

Supervisors were responsible for ensuring that enumerators understood and continually carried out their responsibilities diligently. They provided the necessary logistical support, oversight, carried out random spot checks, and ensured that surveys were conducted in the selected spots without substitution. The supervisors also conducted quality-assurance tests to ensure that the enumerators were collecting valid data. Additionally, they carried out a 10% full audit of all surveys to ensure that the responses were complete and within expected bounds. They also maintained regular communication with the survey coordinator.

2.2.3 Data Quality Control

Training of enumerators was conducted on the 20th of March 2019. The training included the introduction to the goals and objectives of the survey, introduction to the interviewing process, role-playing, planning for the actual data collection and area allocation. Enumerators were also introduced to the electronic data collection method of administering questionnaires using the tablet, as well as the ability to send in responses in real time. For uniformity, they were trained on the requirements of the project, how to ask and answer technical questions, and what to observe. During the training, transportation modalities, target setting, logistics and security considerations were also discussed.

Pre-testing of the questionnaires was carried out after the training sessions. The pre-testing exercise was a key quality assurance measure that provided a clear picture of the relevance of the survey instruments, the accuracy of the data collection tool (inyst) script and questions sequencing. It also helped us to identify sensitivities and challenges that might arise from administering the questionnaires. Feedbacks from the pre-testing exercise informed the revision and finalization of the survey tools.

Given the extensiveness of the survey and the resultant sensitivity to data accuracy, various risks could threaten the quality of data collected. These included interviewing ‘ghost’ respondents, deliberate skipping of questions to shorten the time taken per questionnaire and inaccurate data entry. The questionnaire CAPI script included validation checks and relevant constraints to guarantee accurate data entry. For instance, devising a constraint to ensure that specific questions were asked before a form could be submitted. The pre-testing exercise helped to ensure the relevance of questions asked and the accuracy of the Inyst script. The data was then cleaned and presented in the preferred spreadsheet format after scrutiny.

3. Empirical Analysis

This study relies on primary survey data on street hawkers in Accra sampled by means of a designed questionnaire and interviews. A total of 351 randomly sampled hawkers at 12 sites in Accra with complete information on all variables were reported for analysis on hawkers' willingness to leave (WTL). After a systematic and pragmatic review, the purposeful Wald test endorsed the following subject-specific variables as significant model predictors: 'Age' [the age of hawker], 'Gender' [whether the hawker is a male or female], 'External financial aid' [Do you have any external financial aid?] 'Road description' [Road description at hawking spot], 'Arrested before' [Have you ever been arrested for hawking?], 'Daily sales revenue' [On an average, how much income do you make daily through hawking? (Gh₵)], 'Amount saved monthly' [How much of the income earned through hawking do you save monthly? (Gh₵)] and 'Working experience' [How long have you been hawking on the streets of Accra? (Months)].

Each hawker's response represents a cluster of correlated observations of the repeated outcome variable coded as '0's and '1's where '1' means that the street hawker is willing to exit street hawking assuming the hawker's desired financial expectations towards an alternative business is met, and '0' represents otherwise. Attributable to the challenges associated with contact tracing of hawkers over long periods, the repeated responses were sampled over close time intervals, hence contributing to high positive correlations at all response levels. The randomly sampled hawkers responded to the question on willingness to leave repeatedly on three occasions while hypothetically relaxing relatively to hawkers, the conditions of financial expectations by -5%, -10% and -15% on the first, second and third occasions respectively.

From the data gathered, we realized minimal variations in responses representing the hawkers' firm decision on the choice on willingness to leave street hawking. On the first and second occasions, about 43.3% and 43.0% of the respondents were willing to leave street hawking respectively whilst on the third occasion, the percentage reduced to 41%. Evident in Figure 2, the distribution of hawkers over willingness to leave do not vary greatly across all occasions where the response was measured. Also, the boxplots displayed a right skewed distribution for the four variables across the choices on all occasions where the responses were recorded. Although it may not be obvious in the plot to emphatically suggest that older hawkers are more willing to quit hawking given another alternative, the majority of the hawkers who responded positively were those who averagely on a daily basis made less income and saved on a monthly basis, a negligible portion of their profit. The summary statistics on the continuous predictors are also reported in Table 1. As can be observed from the table, the value of skewness indicates a right skewed distribution which is fairly approaching symmetric for the age of hawkers' variable implying that the majority of the respondents were youth below 24 years. Similarly, a strictly right skewed distribution is perceived for the other continuous variables suggesting that majority of the observed data being less than the observed mean.

Table 1: *Summary statistics of Willingness to leave street hawking and continuous predictors*.

Variable	Mean	Std.Dev.	Min	Median	Max	Skewness	Kurtosis
Age (yrs)	24.114	8.247	8	22	44	0.384	2.472
Daily sales revenue (Gh₵)	50.866	42.310	5	45	500	4.104	38.728
Amount saved monthly (Gh₵)	125.883	137.455	0	100	1000	2.116	11.711
Working experience (months)	18.995	25.761	0.099	12	288	5.382	45.215

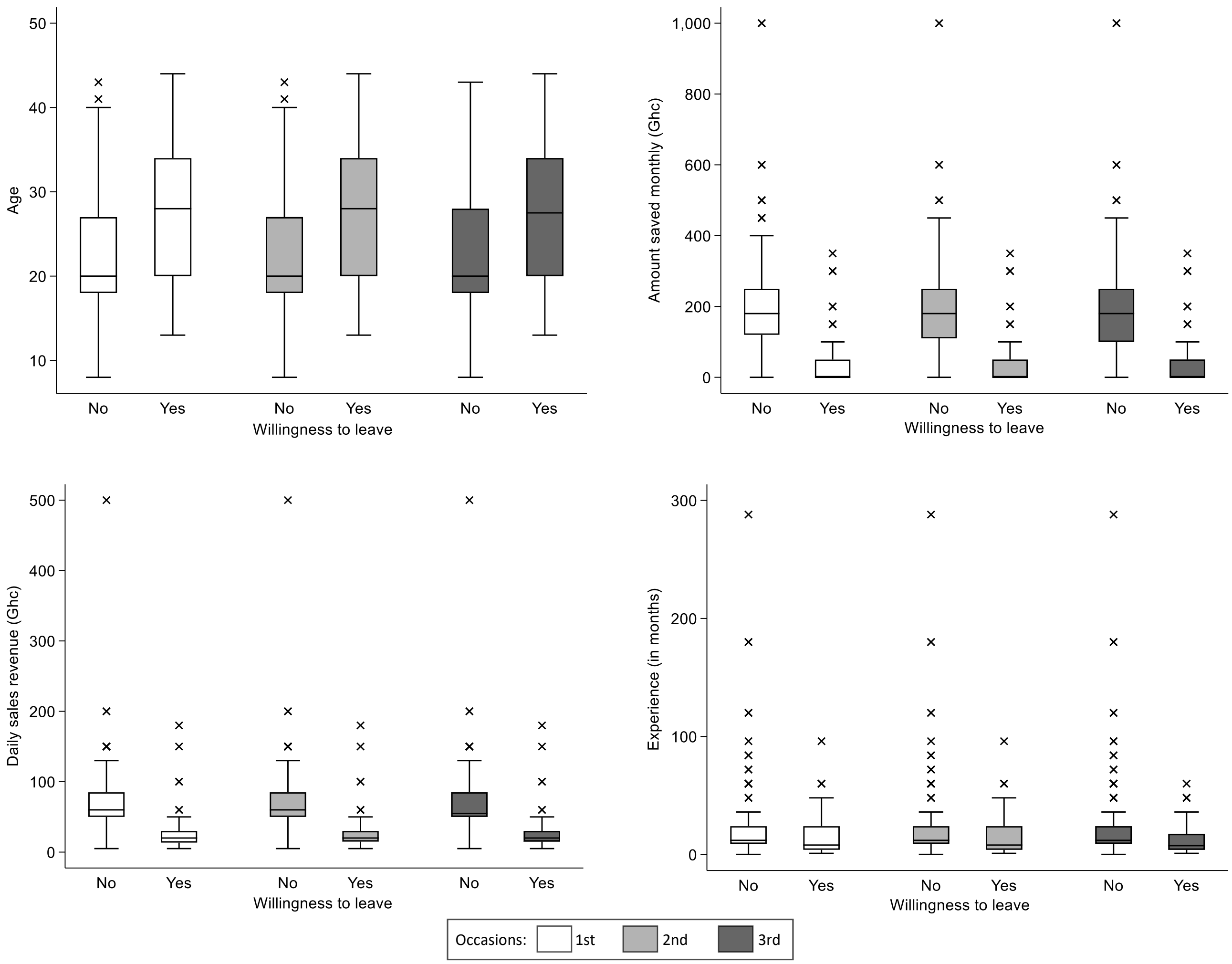

Figure 2: *Boxplot of age of hawker (panel 1), amount saved monthly (panel 2), daily sales revenue (panel 3) and hawker's working experience (panel 4) against willingness to leave street hawking across different time indicators*.

To complement the sampling approach employed, the GEE model is proposed to capture the correlations existing within the dichotomous responses. After a comparison of $QICu$ values displayed in Table A.1 of the Appendix, we endorsed the model with all the eight variables as the best combination of significant covariates. Eventually, the square root transform of the variable 'monthly saving' was considered with the aim of attaining a good model fit while retaining observed zeros which is indicative of hawkers who do not actually save portions of their proceeds.

In Table A.2, we present outputs on a naïve model which assumes universal independence, an unstructured correlation structure model with robust standard errors, an exchangeable correlation model with robust error estimates, and an AR 1 correlation model with robust error estimates. Relatively, the general inferences about the effect of covariates and estimates only deviates marginally amongst the models, portraying universally consistent estimates across the specified models. Moreover, the fact that majority of the within-subject responses remained invariant over time modelled with subject-specific covariates also contributes to the closeness of model estimates.

Without accounting for within correlations, the standard error estimates for the independent model appears intriguing, possessing the least error estimates and a reduced confidence interval. Thus, ignoring the correlation structure leads to the smaller p-values, inconsistent and sometimes misleading inferences. For instance, the variable 'daily sales revenue' is reported as highly significant at a 0.01 level for the naive model but only significant at a 0.1 level for the other robust models except for the unstructured correlation model which is not significant. Hence for the purposes of testing, the estimated standard errors for the naïve model are not considered reliable as compared with the model that account for within-subject correlation though the coefficient estimates may be approximately close.

Coupled with modelling three within-subject responses and the generalization of the covariance structure across subjects, all options of GEE models with covariance structure appears parsimoniously satisfactory in principle. Since there are only three repeated outcomes per subject within the survey design and the order of these responses is key to the choice correlation structure, it is compelling to assume that either the unstructured correlation or the AR 1 correlation structure may be more appealing. Nevertheless, since the within correlation parameters estimated were quite close, an assumption of one scalar parameter as in the common correlation structure might seem reasonable, hence the need for statistical measures to make the correct decision. From Table 2, we observe that though the estimates for the different GEE models may be nearly equal, the unstructured model has the smallest $QICp$ and $QICHH$ values of 654.541 for the GEE logit, thus making the GEE logit model with an unstructured correlation structure more statistically desirable. We also observe an equal $QICp$ and $QICHH$ values suggesting that the coefficient estimates remain unchanged under both the independent model and the specified correlation structure model with robust standard errors.

Table 2: *A comparison of quasi-likelihood information criteria*.

	Logit			
GEE Model	Exch	AR 1	AR 2	Unst
$QICp$	654.735	654.589	654.597	654.541
$QICHH$	654.735	654.589	654.597	654.541

From the final GEE model, the global wild test proved significant with a chi-squared (9 degrees of freedom) statistic of approximately 128.41, thus supporting the claim that the covariates exhibit a substantial effect within the model. Moreover, using four groups, the Horton et al. (1999) chi-squared (3 degrees of freedom) statistic of 3.50 for the final GEE logit model with a p-value of 0.3205 indicate no evidence of lack of fit endorsing the functional form of the model given the sampled data. To add to the argument, we realized that the pseudo R-squared values are all better under the GEE logit link model as compared with the counterpart GEE probit link model further endorsing the GEE logit with an unstructured correlation structure as optimal. It is worth noting that all pseudo R-squared estimates in Table A.3 exceeded 50% with the highest recorded for the Cragg Uhler's statistic at about 74.3% signifying a moderately good model fit.

The final model estimates are presented in Table 3 containing the coefficient and odds ratio estimates with matching robust errors, alongside the p-values and the 95% confidence interval for the odds ratios.

Though the business requires a lot of mobility and energy while competing for market share, the willingness to quit street hawking is not significantly influenced by the hawker's gender as depicted within the model. Likewise, the variable road description with levels, straight roads, and T/X junctions with a roundabout as reference level are statistically insignificant when considering the decision of hawkers to leave street hawking, even though T/X junction and straight roads exhibit a positive and negative effect, respectively, on the log odds of willingness to leave street hawking.

From the output, the respondent's age significantly increases the odds of leaving street hawking for greener pastures by approximately 12%, falling within a 95% confidence interval of (7.3, 17.2) % supposing all other covariates are held constant. This suggests that the higher the age of the hawker, the more likely it is for that individual to opt to leave street hawking.

Assuming that all other predictors are held fixed, the odds ratio of hawkers who have external financial aid relative to those without external financial support in the business is approximately 0.222 yielding a Wald statistic of -3.15 and the corresponding p-value of 0.002 which demonstrates a highly significant effect on willingness to leave street hawking.

Restraining all other explanatory variables constant, the odds of willingness to leave street hawking amongst hawkers who have once been arrested for hawking is nine out of fifty times the odds of street vendors who have never been arrested for hawking. Thus, hawkers who have once been arrested are less willing to leave street hawking, epitomizing their persistence and how they solely rely on proceeds of hawking as life support.

For every unit rise in daily sales revenue made by hawkers, we expect the odds of willingness to leave street hawking to reduce by about 3.6%, assuming all other covariates are held constant though insignificant with a p-value of 0.101.

Table 3: *Willingness to leave street hawking: Parameter estimates (standard errors) for the GEE Logistic model*

WTL		Estimates (s.e)	OR (robust s.e)	p-value	95% C.I for OR
Age		0.115(0.023)	1.122(0.025)	0.000	(1.073, 1.172)
External financial aid	Yes	-1.507(0.478)	0.222(0.106)	0.002	(0.087, 0.566)
Gender	Male	0.309(0.355)	1.362(0.484)	0.385	(0.679, 2.734)
Road description	Straight road	-0.253(0.714)	0.776(0.554)	0.723	(0.192, 3.146)
	T/X-junction	0.533(0.728)	1.704(1.240)	0.464	(0.409, 7.092)
Arrested before	Yes	-1.706(0.547)	0.182(0.099)	0.002	(0.062, 0.530)
Daily sales revenue		-0.033(0.020)	0.967(0.019)	0.101	(0.930, 1.006)
SQRT(Amount saved)		-0.171(0.052)	0.843(0.044)	0.001	(0.761, 0.933)
Working experience		-0.036(0.012)	0.964(0.011)	0.002	(0.942, 0.987)
Constant		0.548(0.857)	1.730(1.482)	0.522	(0.323, 9.275)
Chi-square(9)		128.41			

Likewise, a unit increment in the amount of square root of income saved monthly by hawkers significantly reduces the odds of leaving street hawking by 0.157; thus, the larger the amount saved by hawkers, the less likely it is that they would be willing to leave street hawking.

The coefficient of hawkers' working experience within the model represents an odds ratio of approximately 0.964; hence for every one-month increase in the working experience of street vendors at hawking, we expect a significant decrease of about 3.6% in the odds of willingness to exit street hawking assuming all other variables are kept uniform. This suggests that in comparing hawkers under similar conditions, the individuals with higher working experience are less likely to exit the business.

Conclusions

Street hawking is an essential source of existence for most urban poor in developing countries. While the subject lends itself to rigorous empirical analysis for effective policy recommendations, no such study exists. This paper seeks to fill that gap by using GEE modelling to investigate the willingness of street hawkers in Accra to exit their trade when offered viable alternatives. Assessing street hawkers' willingness to exit hawking will give policy makers flexibility to target multiple segments of street vendors at different incentive levels. Such an analysis could help improve the success of future intervention programmes.

Given that street hawking essentially requires high energy and vitality to excel, the number of hawkers keeps increasing considerably, over time, with the youth constituting the majority group. This study indicates a direct relationship between willingness to leave street hawking and hawker's age. Thus, per the study, the older generation is less likely to survive in the hawking business than the younger generation. In addition, we deduce that the hawkers' willingness to leave may decrease over increasing monthly experience. Thus, a young person who has had several months at hawking is more likely not to entertain the thought of quitting street hawking to start up a new business line soon.

Likewise, we can infer that many hawkers are resilient to the extent that even the fear of being arrested does not deter them from engaging in street hawking. Most hawkers largely hinge on the proceeds of street hawing for their livelihoods. In support, we realize that hawkers who have been arrested before are less likely to leave street hawking since, in most cases, the business serves as life support. Also, from the study, hawkers who received financial assistance to facilitate their business are less likely to exit their trade. Arguably, this might be linked to the fear of losing financial support.

To the hawker, traffic congestion creates a favorable market for hawking and consequently leads to improved sales. As part of the research findings, we observe that increasing sales revenue tends to dwindle the hawkers' willingness to quit street hawking for other business avenues. As well, profit margins and subsequently monthly savings may also rise with increasing sales margin. Plus, the monthly savings made from street hawking may serve as financial security, which may indirectly influence a rejection of the offer to start a new job avenue at the expense of hawking.

Finally, it is revealed that the willingness to quit street hawking is not significantly influenced by the hawker's gender, sales revenue and the road description as depicted within the model.

Appendix A

Table A.1: *Results of $QICu$ values for final covariate selection*

Covariates	$QICu$
WTL on: AGE	1418.44
WTL on: Q31	1388.35
WTL on: GENDER	1646.64
WTL on: LOC_DESC	1625.71
WTL on: Q47	1510.21
WTL on: Q22	1022.37
WTL on: SQRT_Q24	892.176
WTL on: EXPERIENCE	1593.86
WTL on: AGE Q31	1247.49
WTL on: AGE GENDER	1412.54
WTL on: AGE LOC_DESC	1392.92
WTL on: AGE Q47	1301.77
WTL on: AGE Q22	848.644
WTL on: AGE SQRT_Q24	750.018
WTL on: AGE EXPERIENCE	1269.15
WTL on: AGE Q31 GENDER	1245.77
WTL on: AGE Q31 LOC_DESC	1200.70
WTL on: AGE Q31 Q47	1137.21
WTL on: AGE Q31 Q22	822.197
WTL on: AGE Q31 SQRT_Q24	698.897
WTL on: AGE Q31 EXPERIENCE	1134.92
WTL on: AGE Q31 GENDER LOC_DESC	1199.69
WTL on: AGE Q31 GENDER Q47	1132.90
WTL on: AGE Q31 GENDER Q22	823.559
WTL on: AGE Q31 GENDER SQRT_Q24	699.625
WTL on: AGE Q31 GENDER EXPERIENCE	1136.16
WTL on: AGE Q31 GENDER LOC_DESC Q47	1105.27
WTL on: AGE Q31 GENDER LOC_DESC Q22	797.442
WTL on: AGE Q31 GENDER LOC_DESC SQRT_Q24	687.781
WTL on: AGE Q31 GENDER LOC_DESC EXPERIENCE	1105.02
WTL on: AGE Q31 GENDER LOC_DESC Q47 Q22	756.628
WTL on: AGE Q31 GENDER LOC_DESC Q47 SQRT_Q24	655.123
WTL on: AGE Q31 GENDER LOC_DESC Q47 EXPERIENCE	1011.88
WTL on: AGE Q31 GENDER LOC_DESC Q47 Q22 SQRT_Q24	628.459
WTL on: AGE Q31 GENDER LOC_DESC Q47 Q22 EXPERIENCE	707.961
WTL on: AGE Q31 GENDER LOC_DESC Q47 Q22 SQRT_Q24 EXPERIENCE	608.064

		GEE logit Model				
Variables	Levels	ind (s.e)	exch (robust s.e)	AR 1(robust s.e)	AR 2(robust s.e)	unst (robust s.e)
Age		0.113(0.016)***	0.113(0.022)***	0.115(0.023)***	0.113(0.022)***	0.115(0.023)***
External financial aid	Yes	-1.537(0.316)***	-1.537(0.479)***	-1.503(0.477)***	-1.533(0.479)***	-1.507(0.478)***
Gender	Male	0.308(0.222)	0.308(0.357)	0.297(0.355)	0.307(0.356)	0.309(0.355)
Road description	Straight road	-0.241(0.438)	-0.241(0.719)	-0.257(0.714)	-0.243(0.718)	-0.253(0.714)
	T/X-junction	0.532(0.435)	0.532(0.731)	0.523(0.728)	0.531(0.731)	0.533(0.728)
Arrested before	Yes	-1.689(0.318)***	-1.689(0.555)***	-1.717(0.545)***	-1.693(0.553)***	-1.706(0.547)***
Daily sales revenue		-0.034(0.006)***	-0.034(0.020)*	-0.033(0.020)*	-0.033(0.020)*	-0.033(0.020)
SQRT(Amount saved)		-0.174(0.024)***	-0.174(0.052)***	-0.169(0.052)***	-0.173(0.052)***	-0.171(0.052)***
Working experience		-0.033(0.007)***	-0.033(0.011)***	-0.037(0.012)***	-0.033(0.011)***	-0.036(0.012)***
Constant		0.607(0.592)	0.607(0.864)	0.554(0.856)	0.599(0.863)	0.548(0.857)
$\rho_{2,1}$		0.000	0.972	0.984	0.983	0.986
$\rho_{3,1}$		0.000	0.972	0.969	0.980	0.951
$\rho_{3,2}$		0.000	0.972	0.984	0.983	0.954
n		1,053	1,053	1,053	1,053	1,053
m		351	351	351	351	351

Table A.2: *Willingness to leave street hawking: Parameter estimates (standard errors) for the GEE logit model with different 'working' correlation structure*
*** p<0.01, ** p<0.05, * p<0.1

Table A.3: *Pseudo R-square values for the GEE logit and probit models with an unstructured correlation structure*

	unstructured GEE model	
Pseudo R-square	logit	probit
Efron	0.716	0.701
McFadden	0.590	0.583
Ben-Akiva Lerman	0.576	0.569
Cox Snell	0.553	0.548
Cragg Uhler	0.743	0.737

References

Anjaria, J. S. (2006) Street Hawkers and Public Space in Mumbai, *Economic and Political Weekly* 41, 2140-2146.

Aloisio, K.M., Swanson, S. A., Micali, N., Field, A., & Horton, N. J. (2014). Analysis of Partially Observed Clustered Data using Generalized Estimating Equations and Multiple Imputation. *Stata J*. 14(4), 863-883.

Agbo, M. C. (2017). The Health and Educational Concequences of Child Labour in Nigeria. *Health Science Journal* 11(1).

Ballinger, G. A. (2004). Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods*. 7(2): 127-150.

Bhowmik, S. (2000). Street Vendors in Mumbai. *Journal of livelihood* (8), 240-247.

Bhowmik, S. (2003). National Policy for Street Vendors. *Economic and Political Weekly 38(16),* 1543-1546.

Bhowmik, K. S. (2005). Street Vendors in Asia: A Review. *Economic and Political Weekly 40(22-23)*, 2256-2264.

Bhowmik, S (2010a). Legal Protection for Street Vendors. *Economic and Political Weekly 45(51)*, 12-15.

Bhowmik, K. S. (2010b). *Street Vendors in the Global Urban Economy,* 1st ed., Routledge: New Delhi.

Bhowmik, S. (2012). *Street Vendors in the Global Urban Economy,* 1st ed., Routledge: New Delhi.

Bromley, R. (2000). Street vending and public policy: A global review. Intern J Sociol Soc Policy *20*, 1-28.

Bryceson, D., & Potts, D. (2006). African urban economies: Viability, vitality or vitiation? Dev. Matt Chang. 37*,* 1159-1160.

Cox, D. R., & Snell, E. J. (2018). *The analysis of binary data,* 2nd ed., Chapman and Hall: London.

Cragg, J. G., & Uhler, R. S. (1970). The demand for automobiles. *Can. J. Econ. 3*(3), 386-406.

Deroche, C. B. (2015). *Diagnostics and model selection for generalized linear models and generalized estimating equations.* (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/3059.

Dey, D., & Dasgupta, S. (2010). Integration of an informal economy with the globalization process: A study on street hawkers of Kolkata. *Available at SSRN 1584623*. https://dx.doi.org/10.2139/ssrn.1584623

Efron, B. (1978). Regression and ANOVA with zero-one data: measures of residual variation. *J Am. Stat. Assoc.* 73(361), 113-121.

Hardin, J. W., & Hilbe, J. M. (2013). *Generalized Linear Models and Extensions*, 2nd ed. CRC Press: London.

Hin, L. Y., Carey, V. J., & Wang, Y. G. (2007). Criteria for working–correlation–structure selection in GEE: Assessment via simulation. *Am. Stat. 61*(4), 360-364.

Horton, N. J., Bebchuk, J. D., Jones, C. L., Lipsitz, S. R., Catalano, P. J., Zahner, G. E., & Fitzmaurice, G. M. (1999). Goodness-of-fit for GEE: an example with mental health service utilization. *Stat. Med. 18*(2), 213-222.

Hosmer, D. W., & Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model." *Commun Stat. Theory Methods 9*(10), 1043-1069.

Huh, D., Flaherty, B. P., & Simoni, J. M. (2012). Optimizing the analysis of adherence interventions using logistic generalized estimating equations. *AIDS Behav. 16*(2), 422-431.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13-22.

Mathews, G., & Vega, C. A. (2012). *Globalization from Below*: *The world's other economy*, 1st ed., Routledge: London.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*, 2nd ed. Florida: *CRC Press*.

McFadden, D (1974). Conditional logit analysis of qualitative choice behavior. In: Zarembka, P., Ed. *Frontiers in econometrics*, Academic Press, 104-142.

Mitullah, W. V. (2004). A review of street trade in Africa. *Working Draft, Institute for Development Studies University of Nairobi.* Retrieved from http://erepository.uonbi.ac.ke/bitstream/handle/11295/54289/Winnie%20V.%20Mitullah.pdf?sequence=1

Owusu, A. B., Abrokwah, S., & Frimpong, S. (2013). Analysis of the spatial and temporal dynamics of street hawking: A case study of the Accra Metropolitan Area. *J. Geogr. Geol. 5*(4), 169-180.

Panwar, A. M., & Garg, V. (2015). Issues and challenges faced by vendors on urban streets: A case of Sonipat city India. *Int. J. Eng. Technol. Manag. Appl. Sci. Vol. 3*, 1-12.

Pappeswari, C., & Rajalakshmi, S. (2014). Socio-economic conditions of street food vendors-with special reference to Tiruchendur area. *Indian J. Commer. Manag. Stud. 5*(1), 70-77.

Parikh, K. S., Fischer, G., Frohberg, K., & Gulbrandsen, O. (1988). *Towards free trade in agriculture*. Martinus Nijhoff Publishers: Dordrecht.

Pekár, S., & Brabec, M. (2018). Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. *Ethology*, *124*(2), 86-93.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using stata 3rd ed.,* Stata Press: Texas.

Roever, S., & Skinner, C. (2016). Street vendors and cities. *Environ. Urban. 28*(2), 359-374.

Steel, W. F., Ujoranyi, T. D., & Owusu, G. (2014). Why evictions do not deter street traders: Case study in Accra, Ghana. *Ghana Soc. Sci. J. 11*(2), 52-76.

Te Lintelo, D. J. (2010). *Advocacy coalitions influencing informal sector policy: The case of India's national urban street vendors policy*. 1st ed. Routledge: New Delh.

Woods, T. (2013). Letter from Ghana: street hawkers. The Guardian. London. Retrieved from https://www.theguardian.com/world/2013/jun/04/guardian-weekly-letter-from-ghana

Yeboah, T., Owusu, L., Arhin, A. A., & Kumi, E. (2015). Fighting poverty from the street: Perspectives of some female informal sector workers on gendered poverty and livelihood portfolios in Southern Ghana. *J. Econ. Soc. Stud. 5*(1), 239-267.