Journal of Choice Modelling 31 (2019) 86–103

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Arriving at a decision: A semi-parametric approach to institutional birth choice in India

Prateek Bansal*, Ricardo A. Daziano, Naveen Sunder

Cornell University, United States

ARTICLE INFO

JEL classification:
I25
I12
C25
C14

Keywords:
Institutional delivery
India
Discrete choice
Semi-parametric methods
Revealed preferences

ABSTRACT

The Multinomial Logit (MNL) model is popular, but a semi-parametric specification of its link/utility function has seldom been used in empirical applications. This is primarily because of the resource intensive nature of semi-parametric estimation. In this paper we propose and implement a parallel computation algorithm to estimate the semi-parametric kernel MNL model. This algorithm reduces model estimation time by a factor of 2–10, depending on the size of the dataset and the available resources for computation. These computational gains make the estimation of this model feasible for large datasets. Additionally, using a Monte Carlo study we show that the kernel MNL outperforms the traditional linear MNL model in terms of fit and predicted choice probabilities. We demonstrate how kernel-based specification can unearth important heterogeneities in the effect of covariates through an empirical exercise. We use data from a nationally representative household survey (N = 157,804) to analyze the factors associated with institutional births (as opposed to home births) in India. Our revealed-preference results indicate that maternal education, household assets, distance to formal health facility, and birth order play an essential role in determining birth location choice. Although the directions of impact are similar across both the linear and the kernel MNL specifications, there are significant differences in the marginal effects of different factors across the two models. These differences, which arise due to the flexibility afforded by the semi-parametric specification, potentially bring additional nuance to policy discussions.

1. Introduction

In this paper we analyze healthcare demand in the context of India. This is important since a bulk of health interventions have focused on the supply side, i.e. these interventions are aimed at improving the quality of healthcare provided to consumers. Since health outcomes are a result of the interaction between health care demand and supply, a comprehensive understanding of the demand is equally critical for policy analysis. In particular, in this research we examine the factors associated with the decision of pregnant women (and their families) to give birth in an institutional birth facility.

Maternal mortality rates are high, and the various complications that happen during child birth are among its major causes (Bartlett III et al., 1993; Chalumeau et al., 2000; Kusiako et al., 2000). Due to a high mortality concentration around the first 24 h after birth, the proponents of institutional birth argue that appropriate care during this critical time is necessary to minimize the risk of mortality (Organization, 2006). Filippi et al. (2006) posit the role of timely and qualified care in reducing maternal mortality and in enhancing well-being of both mother and child.

Institutional birth has several effects on the mother's and child's well-being. First, child birth in a medical facility greatly reduces

* Corresponding author.
E-mail address: pb422@cornell.edu (P. Bansal).

https://doi.org/10.1016/j.jocm.2019.04.001
Received 4 June 2018; Received in revised form 24 March 2019; Accepted 1 April 2019
Available online 20 April 2019
1755-5345/

the chances of infant and/or maternal mortality (Darmstadt et al., 2005; Campbell et al., 2006, 2016). This effect is important especially because of the high rates of mortality due to preventable causes related to pregnancy and child birth – estimates from UNICEF (2018) indicate that at the global level these account for nearly 15,000 deaths per day. Second, previous studies have shown that giving birth in an institutional facility is positively associated with better post-natal care, such as vaccination and breast-feeding (Biks et al., 2015; Setegn et al., 2011; Odiit and Amuge, 2003), especially in developing country contexts. These in turn have been shown to have long term impacts on child outcomes (Kramer et al., 2008; Oddy et al., 2010; Quigley et al., 2012; Belfield and Kelly, 2012). Third, early childhood (0–5 years) is a key period of life – events and circumstances in this period have a large bearing on the physical and cognitive development of children, and have effects that persist into adulthood, in the form of education and labor market outcomes (Currie and Vogl, 2013, provides a detailed review).

From 1990 to 2013, there have been tremendous improvements in the global rates of institutional birth, but several regions of the world are still lagging behind. The proportion of births attended by a skilled attendant has increased from 57% to 74%, while the percentage of births with at least one ante natal visit has risen from 65% to 83% (WHO, 2015; UN, 2014). Despite these advances, maternal mortality in developing countries is more than 14 times the rate in developed countries (Alkema et al., 2016). Although large parts of Northern Africa and Southern Asia have seen rapid declines in maternal mortality, the absolute levels still remain high.

In this paper, we study the factors affecting demand for institutional birth among households in India. In particular, we focus on the role of four critical factors in shaping institutional birth choice in India – maternal education, distance, wealth and birth order. We choose India because it is the second most populous country in the world and still performs poorly on these indicators. For example, maternal mortality in India has fallen from 556 per 100,000 live births in the early 1990s to around 174 per 100,000 live births in 2015, but is still more than twice the target level of 70 deaths per 100,000 live births set in the Sustainable Development Goals (SDG).

In our analysis, we use a Multinomial Logit (MNL) model, which is a limited dependent variable regression that characterizes the association between various covariates and an unordered categorical response variable (McFadden, 1974). The conventional MNL model can be interpreted as a linear indirect utility specification if alternative-specific attributes are present in the context of choice between differentiated goods. In this paper, we explore non-linear relationships, and show that this “linear MNL” specification may not capture all the complexities of modeling birth location preferences in India.

In applications across different disciplines, previous studies have introduced non-linearity in different ways, for example a logarithmic transformation (Krishnamurthi and Raj, 1988), and a piece-wise linear function (Kalyanaram and Little, 1994), among others. Only a handful of studies have addressed non-linearity by using semi- or non-parametric utility specifications that do not require apriori assumptions on the functional form of the logit link.

Abe (1999) first introduced a spline-based utility specification in a semi-parametric MNL. Kneib et al. (2007) extended this model using penalized B-spline in a Bayesian estimation framework, while (Fukuda and Yai, 2010) used cubic spline functions. Further, Müller (2001) introduced kernel methods for binary logit models in which the likelihood is localized by kernel smoothing, whereas Langrock et al. (2014) generalized this estimator to multinomial unordered response variables. In this study, we build upon Langrock et al. (2014)'s work and estimate birth location preferences using a kernel-based MNL approach.

This study makes several contributions to the literature. First, this is one of the first studies to implement semi-parametric logit specifications of the utility function (as opposed to heterogeneity distributions), which are still uncommon in practice, especially in contexts with large datasets (> 150,000 observations). Their usage has been restricted due to their resource-intensive sequential estimation – which arises because of an associated increase in the number of identified parameters. We disentangle the estimation of the kernel MNL into independent tasks, and implement a parallel estimation strategy which quickens the estimation process by a factor of 2–10, depending on the sample size and available computational resources. Since the semi-parametric specification improves on an extensively used parametric MNL approach, this parallelization technique broadens the use of kernel-based MNL models.

Second, we propose an empirical procedure to analyze the semi-parametric results in comparison with the linear MNL model results. Our results indicate that the kernel MNL model is able to provide insights that the linear MNL model cannot, especially in terms of identifying “non-linearity” in the underlying relationship between the covariates and the outcome (health care facility choice). Third, using a Monte Carlo study, we illustrate the superiority of the kernel MNL model based on different statistical criteria (e.g., model fit and predictive analysis). Fourth, we come full circle and provide a framework to compare the semi-parametric specification with the “nearest” linear specification – the underlying parametric model that most closely mimics the semi-parametric results. In a case study, we back-out an empirical procedure to select an appropriate parametric specification that is “closest” to the results of the kernel MNL model. We also provide a graphical framework to analyze the results from the non-parametric analysis.

The paper is organized as follows: section 2 explores the strands of literature that this study contributes to, while section 3 details the data used in the analysis, section 4 describes our empirical strategy in detail. We outline the simulation strategy and results in section 5, and discuss results of the empirical study in section 6. We end with a summary of our findings, their implication in the policy realm and the potential wider use of this technique to similar research questions.

2. Literature review

In this paper we focus on the decision to seek healthcare and the type of health facility chosen. Toward this end, we specifically examine the role played by maternal education, household assets, distance to facility, and birth order of the child in determining birth facility choice. These factors have been identified based on their relative importance in determining health care choice from previous studies in similar contexts.

The education level of the mother can be vital in driving birth facility choices as it has been shown to have a strong impact on a variety of health outcomes and related behaviors (see Grossman, 2006, for a detailed review). In relation to child birth and associated outcomes, maternal education has a causal impact on infant birth weight and fertility (Currie and Moretti, 2003; Lance et al., 2011; Grytten et al., 2014). Developing country evidence shows that enhanced maternal education is associated with improvements in birth weight (Chou et al., 2010; Güneş, 2015), institutional birth (Feyissa and Genemo, 2014), infant mortality (Grépin and Bharadwaj, 2015; Chou et al., 2010), and fertility/pregnancy rates (Duflo et al., 2015; Baird et al., 2011; Breierova and Duflo, 2004).

We include a measure of household assets as a determinant of birth choice location. In a setting such as India, costs of child delivery in a formal health care facility can potentially be very high, especially for poor and vulnerable populations (Garg and Karan, 2008; Bonu et al., 2009). Such high costs of health care are typically financed through borrowing, which requires collateral (as security), or through the sale of household assets (Skordis-Worrall et al., 2011). Therefore, household assets and wealth provide a measure of the ability of the household to bear these out-of-pocket medical expenditures (Peters, 2002). In this study, we quantify assets using an index that accounts for the presence of 25 different items in the household at the time of the survey – such as electricity, water source, television, tractor, car, and roof quality, among many others.

Distance to healthcare facility is an important determinant of the facility choice, especially in developing countries where transportation costs can represent a high fraction of the total expenses due to the relative difficulty in accessing health facilities (Gething et al., 2012). This is critical since prior literature has indicated that the cost of health care and associated activities –such as travel– have a significant negative impact on the choice of birth facility (Celik and Hotchkiss, 2000; Aseweh Abor et al., 2011; Falkingham, 2003; Bhatia and Cleland, 1995; Navaneetham and Dharmalingam, 2002; Gwatkin et al., 2004; Ladusingh and Holendro Singh, 2007). Elevated transportation costs are likely in large parts of rural India (the context we study), where distances to formal health facilities can be fairly large (MOHFW, 2016). In fact, these distance effects are further accentuated due to low public investment and the relatively poor condition of transportation infrastructure in India (Ronsmans et al., 2006; Sawhney, 1993; Kumar et al., 1997; Chowdhury et al., 2006). Also, distance to a health facility has been shown to have a direct effect on the usage of health services in general, and maternal health care in particular (Elo, 1992; Sawhney, 1993; Hjortsberg, 2003; Sarma, 2009; Kesterton et al., 2010; Kumar et al., 2014).

We also explore the effect of birth order on birth location choice, a relation that has been studied in many different contexts. Using data from thirty two countries from across Asia, Latin America and Sub-Saharan Africa, Guliani et al. (2012) find that birth parity has a negative impact on having an institutional birth. Similarly, Nair et al. (2012) identify negative relationship between the birth order and the odds of an institutional birth in the south Indian state of Andhra Pradesh. This result is similar to findings from other parts of the world, including China (Short and Zhang, 2004), Haiti (Gage and Guirlène Calixte, 2006), Peru (Elo, 1992), Thailand (Raghupathy, 1996), and Vietnam (Toan et al., 1996).

3. Data and context

In our analysis we use the District Level Health Survey (DLHS) in India. This is a household survey conducted by the government of India every five years, with the first round conducted in 1998–99. The survey was started with the purpose of monitoring the progress of the ongoing health and family welfare programs in the country. In addition, the survey was also meant to provide data on the status of maternal and child health, family planning and other reproductive health indicators at regular intervals. In this analysis, we use data from Round 3 (2007–08). This round covered around 600 districts out of a total of around 630 districts in the country at the time. Given the nature of the data collected and the geographical area covered, the survey is extremely suitable for the kind of analysis that we conduct in this paper. The survey provides detailed information on health care usage during child birth for all children born in the 5 years preceding the survey date. This information is available at the mother's level, along with other demographic information. Having said that, there is no information on the price of these health care facilities - something that does not have a major effect on our analysis.[1]

As outlined in the introduction, we focus our analysis on the following explanatory variables: maternal education, household assets, distance to health facility, and birth order of the child. Maternal education is measured as the highest grade attained by the mother. A large number of mothers in our sample have no education, and hence have been assigned a value of zero. Household assets are measured as an index that goes from zero to one - where a higher value indicates a larger asset level. This index was created by the team that collected the data and is based on information collected on various household durables and non-durables present at the time of the survey. The birth order of the child is calculated based on the number of live births that the mother had before that particular child was born.

Distance to formal health facility could be measured in many different ways. The alternative measurements depend on the type of health facility that one would want to consider. At a local level, there are Primary Health Centres (PHC) and Community Health Centres (CHC) that provide health care to people. Typically there is one PHC or CHC for a cluster of villages (depending on the size of the villages). Hence, travel time to these facilities may be short, but the variety of available resources are typically minimal. Another option is the District Hospital (DH), which is a government run health facility that is ubiquitous in each district. This is a larger health facility with a larger set of resources. Although there are large regional differences in their quality, DHs are typically the closest formal facility that would provide a wide suite of medical services, especially in rural areas. With modernization of India, there has been a growth in the number of private health care facilities. However, formal private care still is mostly an urban phenomenon. As per the definition in the DLHS data, these include private providers who are part of the formal health care framework, and do not

[1] This is primarily because of two reasons: (i) Most (more than 60%) births in India happen at home, where the cost of delivery may be zero (or minimal). (ii) A bulk of the health facility choices (modelled here) are run by the government and have low (or zero) user fees. Even in places where user fees maybe high, there are several government policies (like health insurance and other incentives) that subsidize the use of these facilities, especially among the poor. Example - Janani Suraksha Yojana, a healthcare scheme of the government of India.

include the informal health care providers discussed earlier.

For multiple reasons, we measure the distance to formal health care facility as the distance to DH. In comparison to PHC/CHC, DH has more resources and hence would be able to handle the various complications that might occur during child birth. Although some PHC/CHC may have medical professionals who can effectively handle most child birth cases, it may not be true in all regions. Given this regional diversity, we decided to consider distance to DH in our models. The reason distance to private facility is not considered as a covariate is because it is not as widely present as DH and the distance to these private facilities is highly correlated to the distance to DH – since both of them are mostly located in the closest urban/semi-urban location with reference to the surveyed household.[2]

4. Empirical methodology

We model birth location choice using an MNL framework. To do so, we first write a generalized framework for an MNL model which subsumes both parametric and semi-parametric specifications. Consider P_{ki}, the conditional probability of choosing alternative k by individual i, specified as:

$$P_{ki}(\mathbf{\Psi}) = \frac{\exp(U_{ki})}{\sum_{j=1}^{K}\exp(U_{ji})} = \frac{\exp(\mathbf{X}_i^t\boldsymbol{\beta}_k + T_k(\mathbf{Z}_i))}{\sum_{j=1}^{K}\exp(\mathbf{X}_i^t\boldsymbol{\beta}_j + T_j(\mathbf{Z}_i))}, \tag{1}$$

where individual-specific covariate vectors are $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$.[3] We divide the set of individual specific covariates into two categories ($\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$) to explicitly account for the fact that our semi-parametric kernel MNL model can handle linear portions. In our kernel MNL framework, $\boldsymbol{X}_i$ enters the model in a standard linear functional form, whereas $\boldsymbol{Z}_i$ is included in the model using a flexible non-linear specification. In our case, the vector $\boldsymbol{Z}_i$ consists of the following covariates of interest: maternal education, asset index, distance to district hospital and birth order. All other covariates are included in $\boldsymbol{X}_i$. Also, $\mathbf{\Psi}$ is a vector that consists of $\boldsymbol{\beta}_k$ and $T_k(\boldsymbol{Z}_i)$, which themselves are vectors of unknown parameters and functions (that will be estimated). For identification, we consider the K^{th} alternative as the base, i.e. $\boldsymbol{\beta}_K = T_K(\boldsymbol{Z}_i) = 0$.

4.1. Estimation of a linear MNL model

The generalized framework in equation (1) nests the conventional linear MNL model. Assuming $T_k(.)$ to be a linear function of $\boldsymbol{Z}_i$, i.e. $T_k(\boldsymbol{Z}_i) = \boldsymbol{Z}_i^t\boldsymbol{\gamma}_{\boldsymbol{k}}$, our framework obviously simplifies to the linear MNL model. This specification has an identified parameter vector $\mathbf{\Psi} = \{\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{K-1}, \gamma_1, ..., \gamma_{K-1}\}$. If each $\boldsymbol{\beta}_k$ and $\boldsymbol{\gamma}_{\boldsymbol{k}}$ have α and r elements, respectively, then the total number of identified parameters is $(\alpha + r) \times (K - 1)$. These parameters can be estimated by maximizing the sample log-likelihood:

$$\mathscr{L}(\mathbf{\Psi}) = \sum_{i=1}^{N}\ln\left(\prod_{k=1}^{K}(P_{ki})^{d_{ki}}\right) = \sum_{i=1}^{N}\sum_{k=1}^{K} d_{ki}\left[\ln\left(\frac{\exp(\mathbf{X}_i^t\boldsymbol{\beta}_k + \boldsymbol{Z}_i^t\boldsymbol{\gamma}_{\boldsymbol{k}})}{\sum_{j=1}^{K}\exp\left(\mathbf{X}_i^t\boldsymbol{\beta}_j + \boldsymbol{Z}_i^t\boldsymbol{\gamma}_{\boldsymbol{j}}\right)}\right)\right], \tag{2}$$

where d_{ki} is a choice indicator that equals 1 if individual i chooses alternative k. In our case, the alternatives are four, namely home birth, CHC/PHC, private clinics, and district hospital.

4.2. Parallel estimation of kernel MNL model

To better characterize the association of certain covariates with the choice variable, we adopt a semi-parametric kernel approach which affords more flexibility than the linear model described above (Langrock et al., 2014). This general model allows for a subset of covariates ($\boldsymbol{Z}_i$) to enter the structure equation with a non-linear functional form. Since this functional form of $T_k(\boldsymbol{Z}_i)$ is unknown, the function T_k is estimated by kernel smoothing. In this case, the identified parameter matrix $\mathbf{\Psi}$ is given by:

$$\mathbf{\Psi} = \{\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{K-1}, T_{11}, ..., T_{1N}, ..., T_{K-1,1}, ..., T_{K-1,N}\} \tag{3}$$

The kernel MNL model is a semi-parametric specification because the identified parameter space includes a vector of finite cardinality ($\boldsymbol{\beta}_k$) and the non-parametric function ($T_k(.)$) which grows with the number of observations (N). Therefore, the total number of identified parameters in the kernel MNL model is $(\alpha + N) \times (K - 1)$. In our case we have in excess of 450,000 (150,000 data points × (4–1) choices) parameters to estimate for each kernel specification.

Since direct maximization of the loglikelihood function is not tractable, we adopt profile loglikelihood estimation (PLE). This is an iterative procedure that allows for separate estimation of the parametric $\boldsymbol{\beta}_k$ and the non-parametric component $T_k(.)$, after

[2] Note that distance represents only one aspect of access, but ignores others such as terrain and availability of good quality infrastructure. We considered the inclusion of the presence of all weather roads in our model, but it was not significant and was hence omitted. This outcome is not surprising as other studies on Indian data have found similar results (Kesterton et al., 2010).

[3] Alternative-specific intercepts are not included in the specification because they are unidentified, i.e. intercepts are indistinguishable from the vertical location of the non-parametric functions $T_k(.)$. Thus, kernel-MNL nests the linear MNL model (without constants) and will always result in better model fit (i.e., loglikelihood), as also highlighted in the Monte Carlo study (see section 5).

conditioning on each other. Severini and Wong (1992) first considered PLE for partial linear regression models and then subsequent work (Müller, 2001; Langrock et al., 2014) extended this procedure to estimate semi-parametric binary and multinomial models. The PLE procedure is iterative in nature, where one updates $T_k(.)$, $\forall\, k$ for each person and $\boldsymbol{\beta_k}$, $\forall\, k$ in each iteration by conditioning on the parameter values (excluding the one being updated) from the previous iteration. More specifically, conditional on $\underset{\cdot}{\boldsymbol{\beta}}$ and $T_{j,\underset{\cdot}{\boldsymbol{\beta}}}(.)$, $j \neq k$ from the previous iteration, the non-parametric function $T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(.)$ is estimated by maximizing the local/smoothed loglikelihood for each individual.[4] Then conditioning on the $T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(.)$, $\forall\, k$ and $\boldsymbol{\beta_j}$, $j \neq k$ from previous iteration, the profile likelihood is maximized to obtain the estimate of $\boldsymbol{\beta_k}$. The iterative procedure terminates when a norm of the difference of consecutive $\underset{\cdot}{\boldsymbol{\beta}}$ attains a value less than a predefined threshold (we worked with 10^{-3}).

4.2.1. Non-parametric estimation of $T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(.)$

For any individual s with $\mathbf{Z_s} = (Z_{s1} \cdots Z_{sr})$, the nonparametric function $T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})$ is estimated in each iteration by maximizing the following *conditional local loglikelihood* (Langrock et al., 2014):

$$\begin{aligned}
&\mathscr{L}_{local}(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})) = \textstyle\sum_{i=1}^{N} (\det\boldsymbol{H})^{-1}\kappa\ [\boldsymbol{H}^{-1}(\boldsymbol{Z_s} - \boldsymbol{Z_i})]\ell_i[.],\\
&\ell_i[.] = \ell_i[\boldsymbol{U}_i(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s}))], \text{ where}\\
&\boldsymbol{U}_i(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})) = [U_{1i,\ldots,}U_{ki}(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})),\ldots,U_{Ki}], \quad \text{such that}\\
&U_{ki}(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})) = X_i^t\boldsymbol{\beta_k} + T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s}) \quad \text{and} \quad U_{ji} = X_i^t\boldsymbol{\beta_j} + T_{j,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_i}), j \neq k,
\end{aligned} \tag{4}$$

and where κ: $\mathbb{R}^r \to \mathbb{R}$ is an $r-$ dimension kernel. $\ell_i[.]$ is the loglikelihood of the i^{th} individual. Note that $\boldsymbol{U}_i(.)$ is only a function of $T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})$ because $\underset{\cdot}{\boldsymbol{\beta}}$ and $T_{j,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_i})$, $j \neq k$ are treated as fixed. To maximize $\mathscr{L}_{local}(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s}))$ for each alternative (home/CHC/private/distric hospital), we obtain the first order conditions (FOCs) w.r.t. $T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})$:

$$\begin{aligned}
&\frac{\partial\mathscr{L}_{local}(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s}))}{\partial T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})} = \textstyle\sum_{i=1}^{N} (\det\boldsymbol{H})^{-1}\kappa(\boldsymbol{H}^{-1}(\boldsymbol{Z_s} - \boldsymbol{Z_i}))\left(\frac{\partial l_i[\boldsymbol{U}_i(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s}))]}{\partial U_{ki}}\right)\left(\frac{\partial U_{ki}}{\partial T_{k,\underset{\cdot}{\boldsymbol{\beta}}}}\right) = 0\\
&\textstyle\sum_{i=1}^{N} (\det\boldsymbol{H})^{-1}\kappa(\boldsymbol{H}^{-1}(\boldsymbol{Z_s} - \boldsymbol{Z_i}))\left(\frac{\partial\ell_i[\boldsymbol{U}_i(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s}))]}{\partial U_{ki}}\right) = 0.
\end{aligned} \tag{5}$$

4.2.2. Parametric estimation of $\boldsymbol{\beta}_k$

After having $T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(.)$, one can compute the profile loglikelihood to estimate $\boldsymbol{\beta_k}$ (Langrock et al., 2014):

$$\begin{aligned}
&\mathscr{L}_{profile}(\boldsymbol{\beta}_k) = \textstyle\sum_{i=1}^{N} \ell_i(\boldsymbol{U}_i(\boldsymbol{\beta}_k))\\
&\boldsymbol{U}_i(\boldsymbol{\beta_k}) = [U_{1i},\ldots,U_{ki}(\boldsymbol{\beta}_k),\ldots,U_{Ki}], \text{ where}\\
&U_{ki}(\boldsymbol{\beta}_k) = X_i^t\boldsymbol{\beta_k} + T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z}_i) \quad \text{and} \quad U_{ji} = X_i^t\boldsymbol{\beta_j} + T_{j,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_i}), j \neq k.
\end{aligned} \tag{6}$$

Now, $\boldsymbol{U}_i$ is a function of only $\boldsymbol{\beta}_k$ because $T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(.)$ and $\boldsymbol{\beta}_j$, $j \neq k$ are treated as fixed. To update $\boldsymbol{\beta_k}$ for each category, we obtain first the FOCs w.r.t. $\boldsymbol{\beta_k}$:

$$\frac{\partial\mathscr{L}_{profile}(\boldsymbol{\beta}_k)}{\partial\boldsymbol{\beta_k}} = \sum_{i=1}^{N}\frac{\ell_i(\partial\boldsymbol{U}_i(\boldsymbol{\beta_k}))}{\partial U_{ki}}\left(\boldsymbol{X_i} + \frac{\partial T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_i})}{\partial\boldsymbol{\beta_k}}\right) = 0. \tag{7}$$

4.2.3. Supplementary computation

To compute the FOCs of the parametric component, we obtain:

$$T'_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s}) = \frac{\partial T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})}{\partial\boldsymbol{\beta_k}} = -\frac{\sum_{i=1}^{N}(\det\boldsymbol{H})^{-1}\kappa(\boldsymbol{H}^{-1}(\boldsymbol{Z_s} - \boldsymbol{Z_i}))\left(\frac{\partial^2[\ell_i(\boldsymbol{U}_i(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})))]}{\partial U_{ki}^2}\right)\boldsymbol{X_i}}{\sum_{i=1}^{N}(\det\boldsymbol{H})^{-1}\kappa(\boldsymbol{H}^{-1}(\boldsymbol{Z_s} - \boldsymbol{Z_i}))\left(\frac{\partial^2[\ell_i(\boldsymbol{U}_i(T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(\boldsymbol{Z_s})))]}{\partial U_{ki}^2}\right)}. \tag{8}$$

The FOCs of $T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(.)$ and $\boldsymbol{\beta}_k$ require first and second derivatives of $\ell_i(\boldsymbol{U}_i)$ w.r.t. U_{ki}:

$$\begin{aligned}
&\ell_i(\boldsymbol{U_i}) = \textstyle\sum_{k=1}^{K} d_{ki}U_{ki} - \ln\left(\sum_{j=1}^{K}\exp(U_{ji})\right)\\
&\ell'_{ki}(\boldsymbol{U_i}) = \frac{\partial\ell_i(\boldsymbol{U_i})}{\partial U_{ki}} = d_{ki} - \frac{\exp(U_{ki})}{\sum_{j=1}^{K}\exp(U_{ji})}\\
&\ell''_{ki}(\boldsymbol{U_i}) = \frac{\partial^2\ell_i(\boldsymbol{U_i})}{\partial U_{ki}^2} = -\frac{\exp(U_{ki})\cdot\left(\sum_{j=1}^{K}\exp(U_{ji})\right) - [\exp(U_{ki})]^2}{\left(\sum_{j=1}^{K}\exp(U_{ji})\right)^2}.
\end{aligned} \tag{9}$$

[4] In PLE, $T_k(.)$ depends on all $\boldsymbol{\beta_k}$, $\forall\, k$ of the previous iteration and this dependence is notationally illustrated by.$T_{k,\underset{\cdot}{\boldsymbol{\beta}}}(.)$

4.2.4. Solution algorithm

To find the solution to the FOCs, we can make use of a Newton-Raphson-type algorithm (Müller, 2001). Thus, we need to compute the second-order derivatives (SODs) for the local/smoothed loglikelihood (equation (4)) w.r.t. $T_{k,\boldsymbol{\beta}_{\cdot}}(\boldsymbol{Z_s})$ and profile likelihood (equation (6)) w.r.t. $\boldsymbol{\beta_k}$.

First-order derivatives (FODs) are rewritten more concisely after substituting equations (8) and (9) and then SODs are derived by further differentiating the FODs. From an algorithmic perspective, we illustrate computation of FODs and SODs at the M^{th} iteration:

$$\begin{aligned} \mathscr{FOD}^M_{local} &= \sum_{i=1}^{N} (\det \boldsymbol{H})^{-1}\kappa(\boldsymbol{H}^{-1}(\boldsymbol{Z_s}-\boldsymbol{Z_i}))\ell'_{ki}(\boldsymbol{U}_i(T^M_{k,\boldsymbol{\beta}_{\cdot}}(\boldsymbol{Z_s}))) \\ \mathscr{SOD}^M_{local} &= \sum_{i=1}^{N} (\det \boldsymbol{H})^{-1}\kappa(\boldsymbol{H}^{-1}(\boldsymbol{Z_s}-\boldsymbol{Z_i}))\ell''_{ki}(\boldsymbol{U}_i(T^M_{k,\boldsymbol{\beta}_{\cdot}}(\boldsymbol{Z_s}))) \end{aligned} \tag{10}$$

$$\begin{aligned} \mathscr{FOD}^M_{profile} &= \sum_{i=1}^{N} \ell'_{ki}(\boldsymbol{U}_i(\boldsymbol{\beta}^M_{\boldsymbol{k}}))(\boldsymbol{X_i} + T'^M_{k,\boldsymbol{\beta}_{\cdot}}(\boldsymbol{Z_i})) \\ \mathscr{SOD}^M_{local} &= \sum_{i=1}^{N} (\det \boldsymbol{H})^{-1}\kappa(\boldsymbol{H}^{-1}(\boldsymbol{Z_s}-\boldsymbol{Z_i}))\ell''_{ki}(\boldsymbol{U}_i(T^M_{k,\boldsymbol{\beta}_{\cdot}}(\boldsymbol{Z_s}))) \end{aligned} \tag{11}$$

Whereas the kernel-MNL is a flexible specification, the estimation procedure turns out to be computationally infeasible for large datasets due to the increase in identified parameters with the number of observations. In our empirical application there are around 450,000 parameters to estimate. However, we noticed that the update of the non-parametric function in each iteration is the most intensive part of estimation; but, this task consists of as many independent operations as the number of observations. This independence encouraged us to parallelize these tasks. Similarly, the parametric vector updates consist of as many independent operations as the number of alternatives. Thus, a similar parallelization strategy can also be employed to update the parametric vector, but computational gains in this dimension will only happen in the case of datasets with large choice sets.

We implement parallelization on a local machine in R (Analytics and Weston, 2014). We efficiently use the multicore functionality of the local machine, instead of using a cluster of computers. As illustrated in Algorithm 1, we distribute independent updates of the non-parametric function and parametric vector on multiple cores of a local machine and collect updated values when tasks on all cores are completed. The simplicity of such parallelization lies in its versatility as it can be adopted by any user with a multicore processor; however, the gain in the computation speed will depend on the number of cores in the local machine. In our case, parallelization could reduce computation time by a factor of ten using twenty-four cores.

Algorithm 1. Parallel Estimation of Kernel-MNL Model.

Algorithm 1: Parallel Estimation of Kernel-MNL Model

Initialization: Initialize parameters at iteration $M = 0$

$\boldsymbol{\Psi}^{(0)} = \{\beta^0_k, T^0_{k,\beta_{\cdot}}(\boldsymbol{Z}_s)\}, \quad k \in \{1,\cdots,K-1\}, s \in \{1,\cdots,N\}$

while $\|\beta^{M+1}_{\cdot} - \beta^{M}_{\cdot}\|_\infty < 10^{-3}$ **do**

Start parallelization: distribute ***Non-parametric function update*** *on multiple cores*

1: **for** $s \in \{1,\cdots,N\}$ **do**

2: **for** $k \in \{1,\ldots,K-1\}$ **do**

3: $T^{(M+1)}_{k,\beta_{\cdot}}(\boldsymbol{Z}_s) = T^{(M)}_{k,\beta_{\cdot}}(\boldsymbol{Z}_s) - \left(\mathcal{SOD}^M_{local}\right)^{-1}\left(\mathcal{FOD}^M_{local}\right)$

4: **end for**

5: **end for**

End parallelization: Collect $T^{(M+1)}_{k,\beta_{\cdot}}(.)$ *from all cores*

Start parallelization: Distribute ***parametric vector update*** *on multiple cores*

1: **for** $k \in \{1,\ldots,K-1\}$ **do**

2: $\beta^{(M+1)}_{\boldsymbol{k}} = \beta^M_{\boldsymbol{k}} - \left(\mathcal{SOD}^M_{profile}\right)^{-1}\left(\mathcal{FOD}^M_{profile}\right)$

3: **end for**

End parallelization: Collect $\beta^{(M+1)}_{\cdot}$ *from all cores*

end

4.2.5. Bandwidth selection

Bandwidth selection in kernel methods is basically the same as model selection. The literature offers various methods for bandwidth selection (Köhler et al., 2014), but cross-validation appears as the most popular and intuitive. In cross-validation, one generates calibration and validation samples by splitting the original sample in two parts. We created Q random partitions such that the *calibration* sample has 95% of the original sample and the remaining 5% is treated as the *validation* sample. We then prespecify the bandwidth grid. For each of the Q partitions and for each bandwidth on the prespecified grid, we estimate the kernel MNL model using the calibration sample, and then obtain the loglikelihood of the validation sample under the model fitted in the estimation stage. Among all given bandwidths, the bandwidth with the highest average loglikelihood of the validation sample over the Q partitions is chosen.

For multidimensional kernels (i.e., $r > 1$), each point on the prespecified grid would be a matrix of $r \times r$. Even if off-diagonal entries of the bandwidth-matrix are assumed to be zero, the number of points on the grid to sufficiently explore the bandwidth space would explode with an increase in the kernel dimension. For example, for a four-dimensional kernel with a diagonal bandwidth matrix, even 10 candidate realizations of bandwidth in each dimension would yield 10^4 points on the grid. For each of 10^4 candidate bandwidth diagonal metrics, Q (say $Q = 25$) models need to be estimated and validated to find the best bandwidth. Since this task quickly becomes computationally intractable, we reparameterize the diagonal bandwidth matrix as $[f\sigma_1, \cdots, f\sigma_j, \cdots, f\sigma_r]$, where σ_j is the standard deviation of the covariate in the j^{th} dimension and f is a scalar multiplier (Langrock et al., 2014). Thus, we need to choose the optimal scalar f with a prespecified grid $f = [.5, .55, .60, ..., 1.40, 1.45, 1.5]$ in this study.

Post-estimation empirical strategies to compute marginal effects and to back-out functional form of the non-parametric component $T_k(\boldsymbol{Z}_i)$ are illustrated when discussing results (Section 6).

5. Monte Carlo study

The kernel MNL is expected to be preferred over linear MNL due to its functional form flexibility, which might enable it to represent the true data generating process more closely. To provide empirical support to this hypothesis, we first conduct a Monte Carlo study where we compare these two empirical frameworks on different statistical criteria.

In the simulated data generating process (DGP), the link function U_{ki} when individual i chooses alternative k is:

$$U_{ki} = \mathbf{X}_i^t \boldsymbol{\beta}_k + T_k(\mathbf{Z}_i) = [X_{i1}\ X_{i2}] \begin{bmatrix} \beta_{k1} \\ \beta_{k2} \end{bmatrix} + T_k(Z_{i1}, Z_{i2}) + \varepsilon_{ki}$$
$$U_{ki} = \beta_{k1} X_{i1} + \beta_{k2} X_{i2} + \underbrace{\rho_{k1} Z_{i1} + \rho_{k2} Z_{i1}^2 + \rho_{k3} Z_{i1}^3 + \rho_{k4} e^{Z_{i2}} + \rho_{k5} Z_{i1} e^{Z_{i2}} + \rho_{k6} Z_{i1}^2 e^{Z_{i2}} + \rho_{k7} Z_{i1}^3 e^{Z_{i2}}}_{T_k(Z_{i1}, Z_{i2})} + \varepsilon_{ki}, \tag{12}$$

where ε_{ki} is a preference shock which is independent across individuals and choices, and is identically distributed Type-I Extreme Value. To simplify MNL estimation for the true specification, the non-parametric segment of the link function $T_k(\mathbf{Z}_i)$ is assumed to be only nonlinear in the covariates Z_{i1} and Z_{i2}, but linear in the parameters $\boldsymbol{\rho_k}$. The attributes Z_{i1} and Z_{i2} were independently drawn from a standard normal distribution. To emphasize on the non-parametric segment of the link function, the covariates in the parametric segment (X_{i1} and X_{i2}) are considered as indicator variables.

From the above DGP, we generated 100 datasets (with 2000 respondents each) with different starting seeds, and each respondent chooses an alternative from a set of three choices. For each dataset, we estimated four MNL specifications: the true specification, linear MNL (with constants), linear MNL (without constants), and kernel MNL.

To evaluate the performance of the kernel model vis-a-vis the linear MNL model, we use statistical criteria that compare the model estimates with those from the true DGP. The four scores that we use are: hit rate, logarithmic score, Brier score, and the spherical score (Kneib et al., 2007). The model with higher score values is preferred. The logarithmic score (or the loglikelihood of the model) is a commonly used criterion to compare models, but is often criticized because it only considers predicted probability of the chosen alternative. Brier and the spherical scores overcome this challenge – they use the entire predictive distribution.[5] Whereas the Brier score creates a true distribution by assigning a point mass of 1 in the chosen alternative and compare it with the predictive distribution, the latter (spherical score) relates the probability of chosen alternative with the norm of the predictive distribution.

The results from the simulation study are presented in the two panels of Table 2 - the top panel presents the average score statistic for each criterion across 100 repetitions of the four models. The bottom panel presents the results on the percentage of the 100 repetitions in which the kernel model outperforms the other models on the different score criteria.

The results indicate that the kernel model outperforms both versions of the linear MNL model (with and without constant). As explained earlier, the kernel model nests the lineal MNL model (without constant) – which is borne out by the simulation results. The bottom panel shows that the kernel model outperforms the linear MNL (without constant) every time on the log-likelihood based scores. More significantly, based on the same criterion the kernel model outperforms the linear MNL model (with constant) around 94 percent of the time, showing that the kernel MNL model may be preferred to the linear MNL model (with constant) that is most commonly used in empirical applications.

In addition to the score criteria (used above), we plot the probability of choosing an alternative under different specifications in all the simulation repetitions to see how well linear and kernel MNL mimic the true DGP. The results of two such iterations are illustrated in Figs. 1 and 2. The kernel MNL is able to retrieve the non-linearities in the data that its linear counterpart could not. This supports the findings from the comparisons made on the basis of the score criteria. Additionally, the non-linearities in the marginal effects using the kernel MNL seem to be a potentially more accurate representation of the true DGP, which exemplifies its importance from an empirical perspective.

[5] The predictive distribution is a vector of the predicated probabilities for each alternative, which is computed using the model estimates.

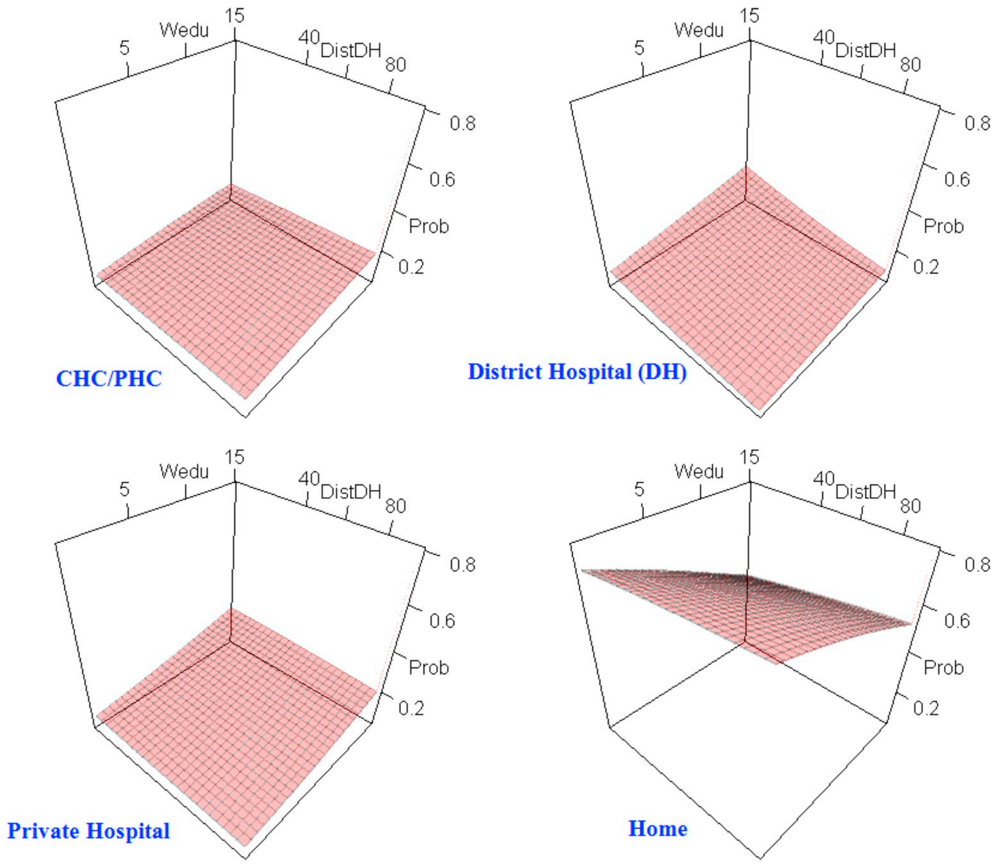


Fig. 1. Simulation accuracy - graph 1.

6. Results – birth location choice

We combine information from multiple survey modules to determine the exact choice set of each individual. Each woman (mother) could have access to a maximum of four choices of birth location - home birth, CHC/PHC, Private Hospital and District Hospital (DH). Table 1 provides summary statistics on related variables. From the table it is evident that home births dominate over other birth facilities with 62.9% share of all births.

We begin our analysis with the conventional linear MNL model, of which we estimate two specifications – both control for the full set of covariates, but only one of them includes state fixed effects (FE). The results are presented in Table 3. We use the linear MNL results (with the base category "home births") as a first step to identifying the covariates that we shall be focusing on in our semi-parametric analysis. In line with the literature on this topic (discussed in earlier sections), we find that the four main continuous variables with significant effects on birth location choice are mother's education, distance to DH, household asset index and birth order. These covariates are all statistically significant at the one percent level in both specifications.

We further estimate a kernel MNL model using the same set of covariates (and the state FE). Whereas we model all key continuous variables non-parametrically, the remaining variables (including state FE) enter linearly in the specification. The local likelihood function uses a Gaussian kernel and bandwidths are selected using a cross-validation procedure. The final results are obtained using a bandwidth of 0.75 times the standard deviation of the continuous covariates.

We first conducted a controlled study to understand a role communication overhead in the parallelization. We took random subsamples of 2,000, 5,000, 20,000, and 150,000 respondents from the original sample and implemented parallel estimation of kernel MNL on 2, 4, 8, and 12 CPU workers. The results are shown in Table 4. Clearly, a high number of CPU workers considerably improves the computational efficiency for large datasets. For example, increasing the number of workers from 8 to 12 reduces the estimation time by a factor of 1.07 and 1.63 for sample sizes of 2 000 and 20,000 respondents, respectively. This indicates that the computational gains in small samples are compensated by the communication overhead and thus, parallelization is more appealing for large datasets where computation times can be reduced by more than a factor of 10.

The MNL model parameter estimates are not readily interpretable – for instance, parameters in a linear MNL model represent marginal utilities in the context of choice among differentiated goods. Therefore, in order to provide a clear interpretation of the kernel-based estimates we plot the probability of choosing an alternative (birth location) against the level of the covariates. To make the impacts of the covariates clearer, we then provide the corresponding (probability) marginal effects at different values of the

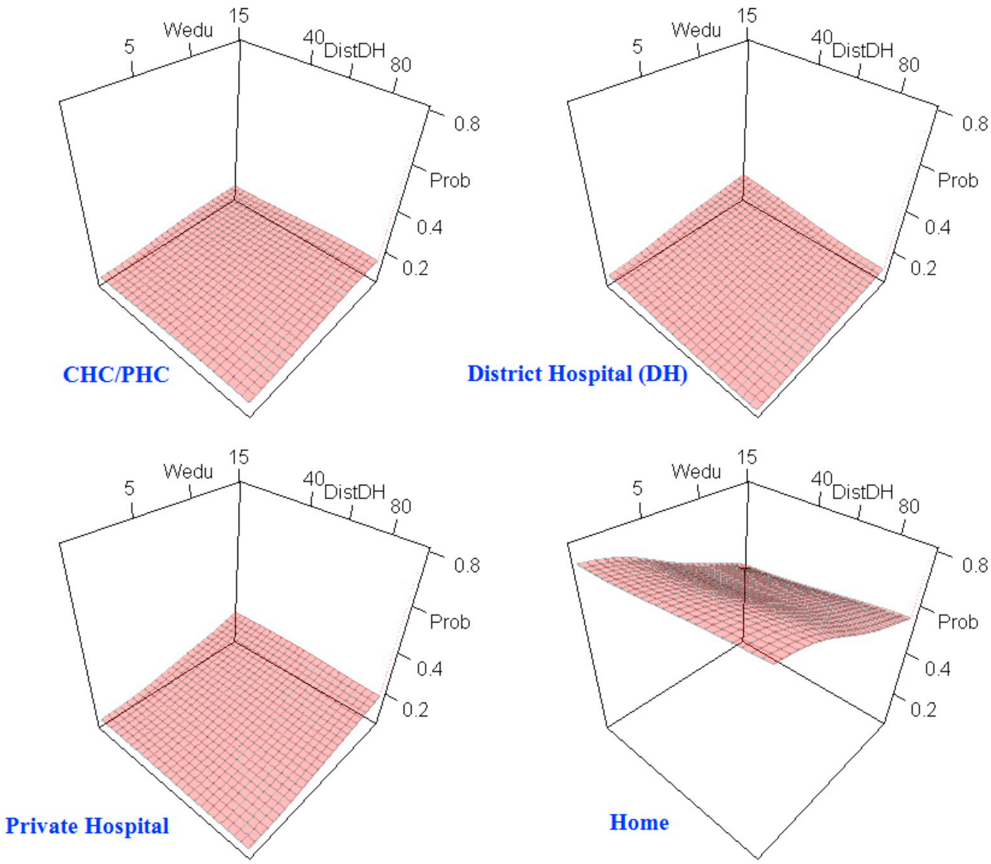


Fig. 2. Simulation accuracy - graph 2.

Table 1
Summary statistics.

	Obs	Mean	Std. Dev.	Min	Max
CHC/PHC	157804	0.11	0.31	0.00	1.00
DH	157804	0.12	0.32	0.00	1.00
Private	157804	0.15	0.35	0.00	1.00
Home	157804	0.63	0.48	0.00	1.00
Muslim Dummy	157804	0.12	0.32	0.00	1.00
SC-ST Dummy	157804	0.40	0.49	0.00	1.00
Female Child dummy	157804	0.46	0.50	0.00	1.00
BPL card dummy	157804	0.34	0.47	0.00	1.00
Household Size	157804	7.00	3.27	1.00	48.00
Asset Index (0–1)	157804	0.24	0.16	0.00	0.96
Birth Order	157804	2.89	1.94	0.00	16.00
Distance to DH	157804	34.99	24.70	0.00	98.00
Mother Education (years)	157804	3.75	4.45	0.00	21.00

covariates.

Since kernel MNL accounts for flexible interaction effects of the covariates entering non-parametrically, variations in the probability of choosing an alternative should ideally be analyzed relative to the value of all the covariates simultaneously. However, for ease of visualization, we restrict ourselves to 3-D probability graphs (cf. Langrock et al., 2014) for linear and kernel MNL, which describe the relation between the probability of selecting an option and its association with the level of two covariates (see Figs. 3–6), while all other covariates are set at their mean values.

Although these 3-D graphs highlight some of the differences in the results across the linear and kernel MNL models, such differences can sometimes be challenging to visualize and interpret. To make these differences more apparent, we derive 2-D plots corresponding to these 3-D graphs, where instead of plotting the probability of selecting a certain birth choice location against two covariates at the same time (3-D graph), we plot this probability against the level of one covariate at a time, while holding all other covariates at their mean (2-D graph). Since our setup has four alternatives, it would be intuitive to have four such plots - each

Table 2
Results of the Simulation study.

Model Specification	Average Scores over 100 iterations			
	Hit Rate	Log Likelihood	Brier	Spherical
True Specification	0.602	−1721.388	−1023.199	1385.711
Linear MNL (with constant)	0.599	−1738.203	−1032.193	1379.884
Linear MNL (w/o constant)	0.584	−1781.808	−1059.605	1360.73
Kernel MNL Model	0.601	−1729.232	−1026.995	1383.535
	Winning % of Kernel model against			
Model Specification	Hit Rate	Log Likelihood	Brier	Spherical
True Specification	43.8	14.6	18	21.3
Linear MNL (with constant)	68.5	94.4	93.3	94.4
Linear MNL (without constant)	85.4	100	100	100

Note: The true data generating process in this case has been assumed to have combination of polynomial and exponential transformation of covariates.

Table 3
Linear MNL models - coefficients.

VARIABLES	Without State FE			With State FE		
	CHC/PHC	DH	Private	CHC/PHC	DH	Private
Mother Education	0.067***	0.118***	0.112***	0.069***	0.093***	0.103***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Asset Index	1.012***	1.987***	3.535***	1.118***	1.676***	3.591***
	(0.067)	(0.064)	(0.058)	(0.073)	(0.073)	(0.066)
Birth Order	−0.200***	−0.273***	−0.217***	−0.161***	−0.234***	−0.198***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)
DH Distance	0.002***	−0.011***	−0.003***	0.001*	−0.013***	−0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Muslim Dummy	−0.427***	0.171***	0.061**	−0.337***	−0.293***	0.062**
	(0.032)	(0.028)	(0.025)	(0.036)	(0.035)	(0.029)
SC-ST Dummy	−0.105***	0.039**	−0.657***	−0.148***	−0.236***	−0.544***
	(0.018)	(0.018)	(0.019)	(0.019)	(0.021)	(0.021)
Female Child	−0.019	0.003	−0.060***	−0.012	0.002	−0.069***
	(0.017)	(0.017)	(0.016)	(0.017)	(0.018)	(0.017)
BPL Card	0.172***	0.187***	0.030*	0.089***	0.051**	−0.159***
	(0.018)	(0.018)	(0.018)	(0.019)	(0.020)	(0.020)
HH Size	−0.009***	−0.046***	−0.021***	−0.004	−0.010***	−0.016***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
Observations	157,804			157,804		
State FE	No			Yes		

Note: The coefficients come from a linear MNL model where the base outcome is home delivery. All coefficients are with respect to that. CHC/PHC stand for Community Helath Centre/Public Health Centre, DH stands for District Hospital and Private stands for private clinics.

Table 4
Computation time of kernel MNL (in hours).

		Cores				Ratio of time (2 and 12 cores)
		2	4	8	12	
Observation	2000	0.131	0.072	0.049	0.046	2.83
	5 000	0.641	0.386	0.240	0.186	3.44
	20000	14.02	7.06	3.96	2.42	5.79
	150000	687.19	–	–	67.02	10.25

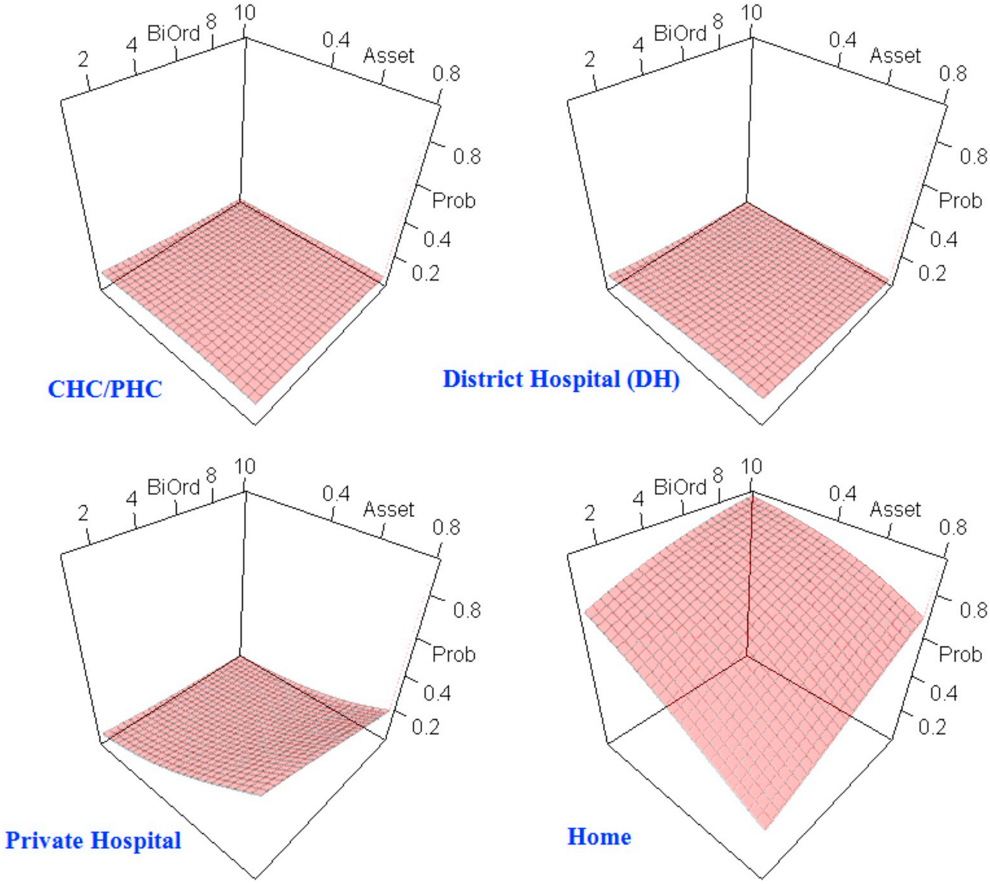


Fig. 3. MNL graph - mother Edu & Dist DH

demonstrating how the probability of choosing an alternative is impacted by a change in the covariate under consideration. Using the fact that the sum of the probabilities across all alternatives is constant (one), we combine these four plots into a single stacked graph (Fig. 7a). This stacked graph shows how the probabilities of each alternative changes as a function of the covariate value, making it easier to observe the co-movements of the probabilities across different alternatives. For instance, Fig. 7a shows that an increase in maternal education leads to a fall in probability of home birth, with a simultaneous rise in the probability of all the other alternatives (Private, District hospital and CHC/PHC).

6.1. Comparison across linear and kernel (2-D probability graphs)

As would be expected, with a rise in maternal education both specifications (linear and kernel MNL) show a decrease in the likelihood of home birth. However, the effects of a change in maternal education in the kernel model are not uniform at different parts of the distribution (see stacked probability plot of kernel MNL in Fig. 7a). To further illustrate the difference between the results from both models, we compare the marginal effects.[6] Instead of point estimates at certain values of the covariate (mean/median/mode among others), we provide the marginal effects at *all* points of the covariate (mother education) distribution. This detailed representation confirms that the difference between the models is a *global* phenomenon and not restricted to certain parts of the distribution.

Fig. 7c plots the marginal effects of mother education on the different birth choice location alternatives, holding all other covariates at their mean level. We find that the linear model might overestimate (or underestimate) the marginal effect on the probability of selecting a certain choice. In our case, the probability of selecting CHC/PHC (Fig. 7c) was overestimated (underestimated) by the linear MNL model at the extremes (middle) of the mother's education distribution. The CHC/PHC subplot of Fig. 7c provides evidence that the linear MNL model implies a relatively flat marginal effect (of around 0.004) until around eight years of education, beyond which there is a fall in its magnitude. This pattern is different from the kernel model where the marginal effect has an inverted U-shape. For women with zero to three years of education, the marginal effect based on the kernel model is less than that of the linear

[6] The marginal effects are the slope of the function that maps the probability of choosing an alternative against a covariate, holding all other covariates at their mean.

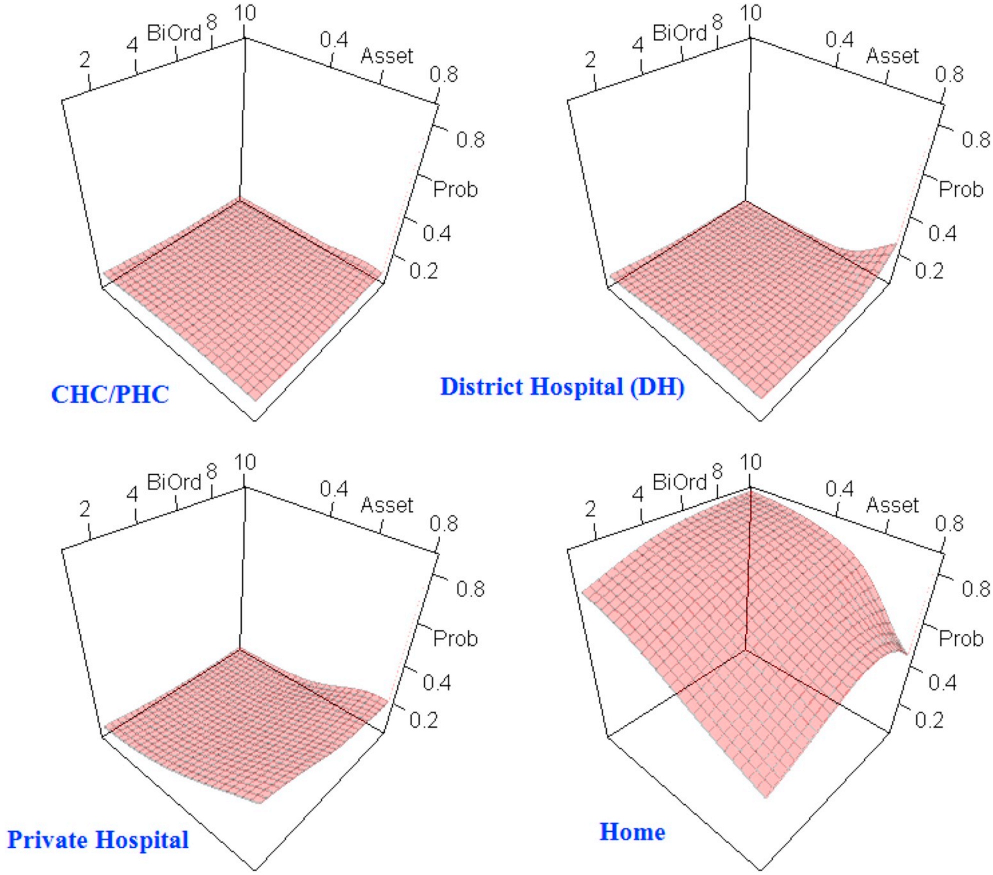


Fig. 4. Kernel graph - mother Edu & Dist DH

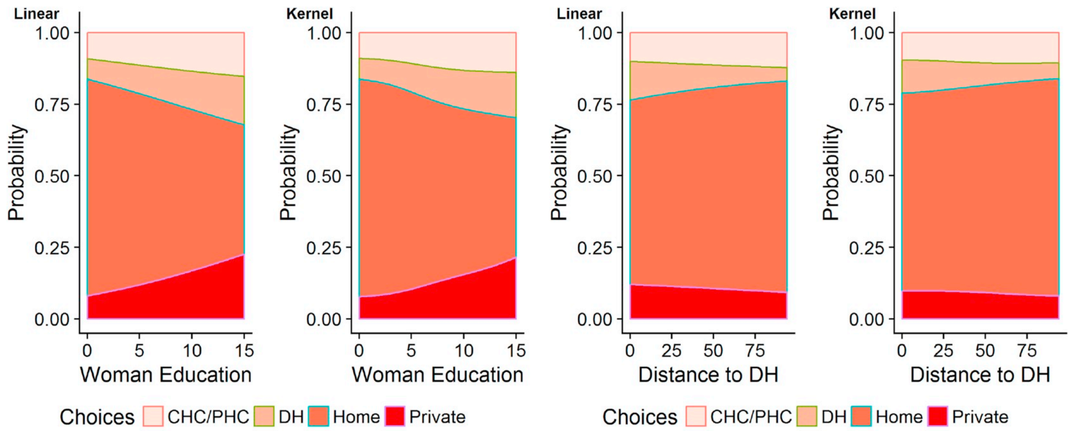
model. The kernel marginal effect rises until around five years of education and then falls to become lower than the linear MNL marginal effect at upper primary school years (grades 7–8 in India). The difference between the two marginal effect estimates is large and is illustrated by the fact that the peak of the kernel marginal effect (around 0.006) is nearly 50 percent higher than the marginal effect using the linear MNL model (nearly 0.004). Additionally, the absolute value of the slope (i.e., the rate of rise and fall) of the marginal effect from the kernel specification is greater than that of the linear specification.

A similar pattern is observed for the choice of district hospital (DH subplot of Fig. 7c). Both the kernel and the linear MNL model have positive marginal effects, i.e. a rise in maternal education always leads to a rise in the probability of choosing district hospital. Here again, the linear model overestimates the marginal effect at the lower and higher ends of the education distribution. As in the previous case, the peak of the kernel model marginal effect is more than 33 percent higher than the corresponding marginal effect for the linear model. Analogously, the marginal effect plots for the cases of private hospital and home births also bear out a similar intuition.

Put together, these results imply that the marginal impact of mother's education on institutional birth choice is not as homogeneous as suggested by the linear MNL specification. The kernel model, through its flexibility, is able to uncover these heterogeneous effects. This variation in the effect sizes based on the level of mother's education is the kind of difference we highlight as an advantage of the kernel MNL model over the linear MNL framework. We emphasize the potential policy relevance of these effect size differences, especially in resource-constrained environments.

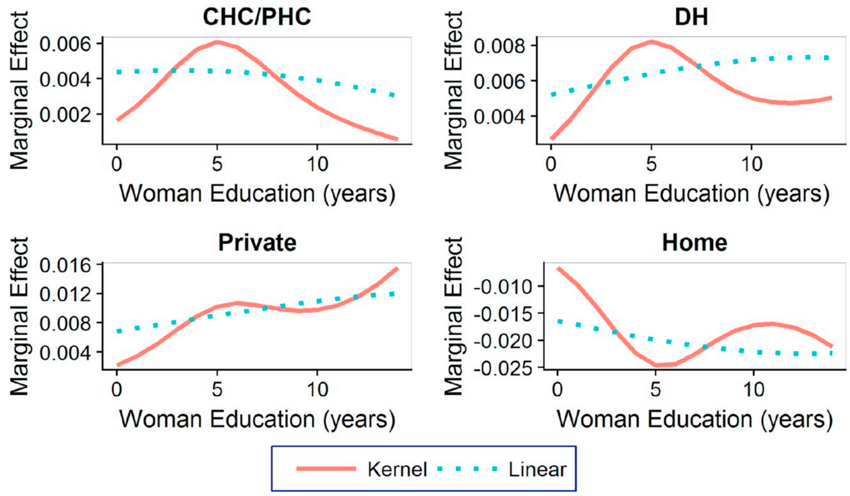
We also examine the impact that household assets have on birth location choice. The Private subplot of Fig. 8c-asset demonstrates that at all points the linear model overestimates the marginal effect of assets on the probability of choosing private hospital for child birth. Given the relatively large magnitudes of the marginal effects, these differences are significant. In a similar vein, although the Home subplot of the same figure (Fig. 8c-asset shows that an increase in assets has a negative (marginal) effect on the probability of home birth in both models, the linear MNL model overestimates the (negative) marginal effect. Similar differences can also be seen in the marginal impact of the asset index under the kernel and linear MNL specifications in the cases of District Hospital and CHC/PHC choices in Fig. 8c-asset.

In our results, we find that although the linear MNL model provides the correct direction of impact, it does not capture critical non-linearities in these relationships. In Fig. 7d the marginal effect of distance to District Hospital on the choice of birth location either has a U-shaped curve (DH and Private) or an inverse U-shaped curve (Home and CHC/PHC). In all the cases the kernel model marginal effects are different from the linear model marginal effects at the extremes of the DH distance distribution (low and high

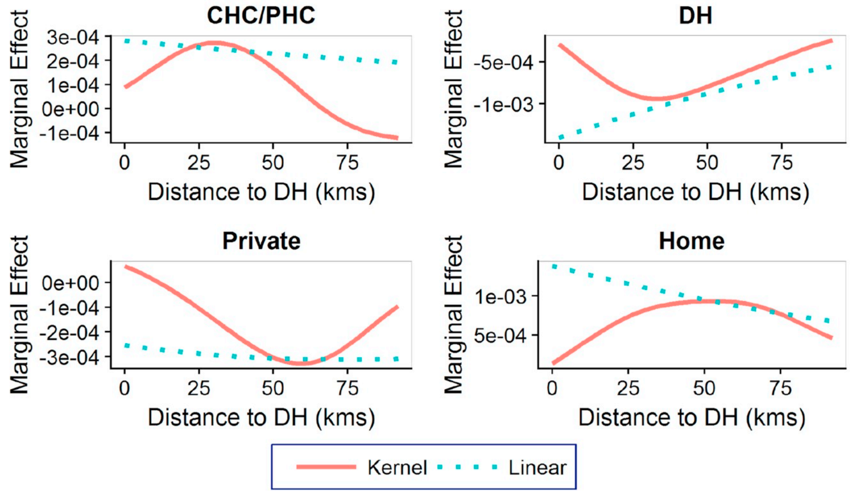


(a) **Impact of Mother edu (others at mean)**

(b) **Impact of Dist DH (others at mean)**

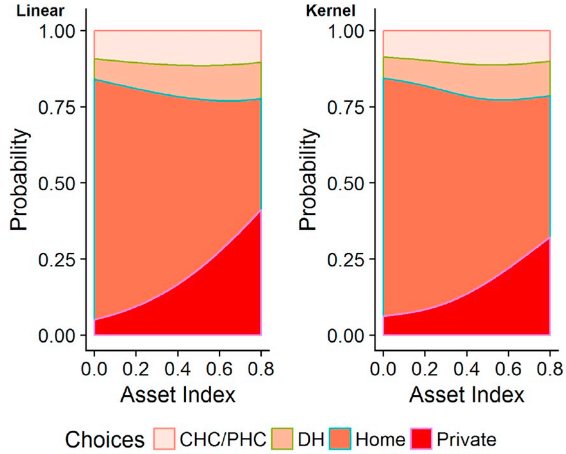


(c) **Marginal Effect - Mother Education**

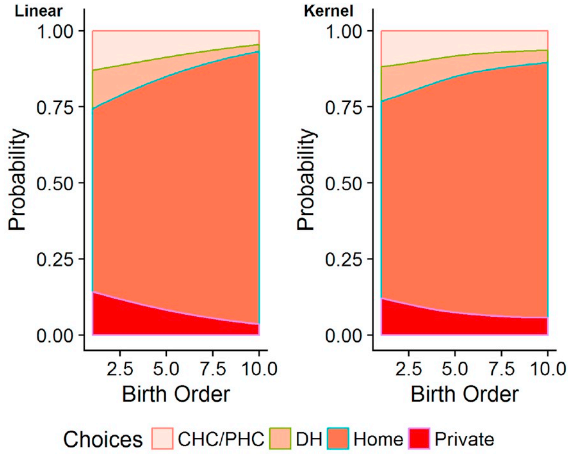


(d) **Marginal Effect - Dist DH**

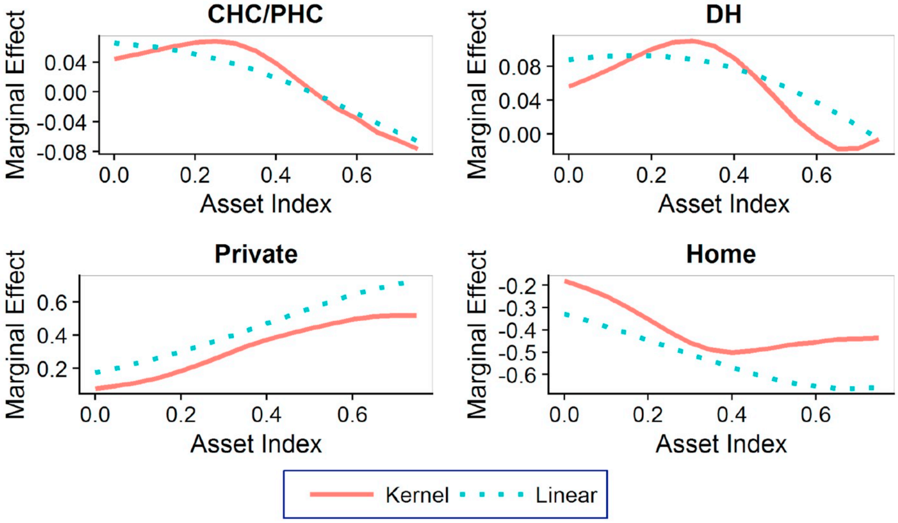
Fig. 5. MNL graph - birth order & asset index.



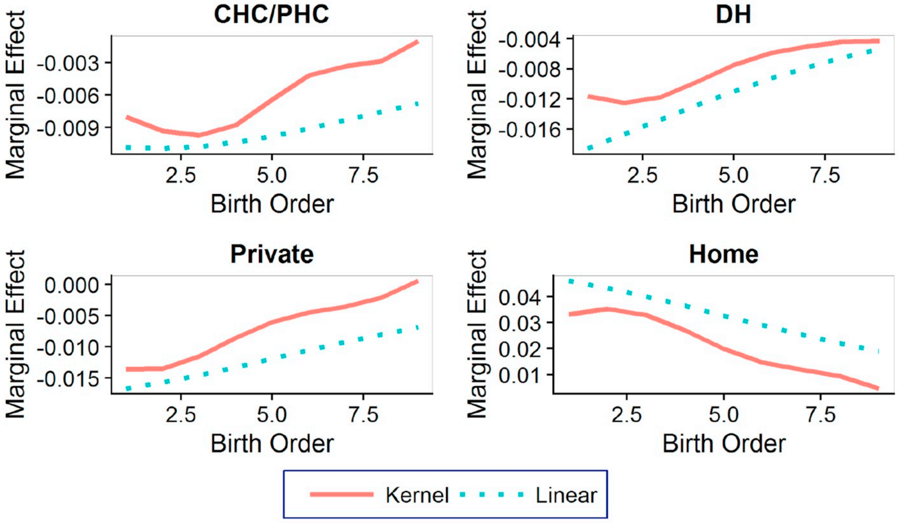
(a) **Impact of Asset Index (others at mean)**



(b) **Impact of Birth Order (others at mean)**



(c) **Marginal Effect - Asset Index**



(d) Marginal Effect - Birth Order

Fig. 6. Kernel graph - birth order & asset index.

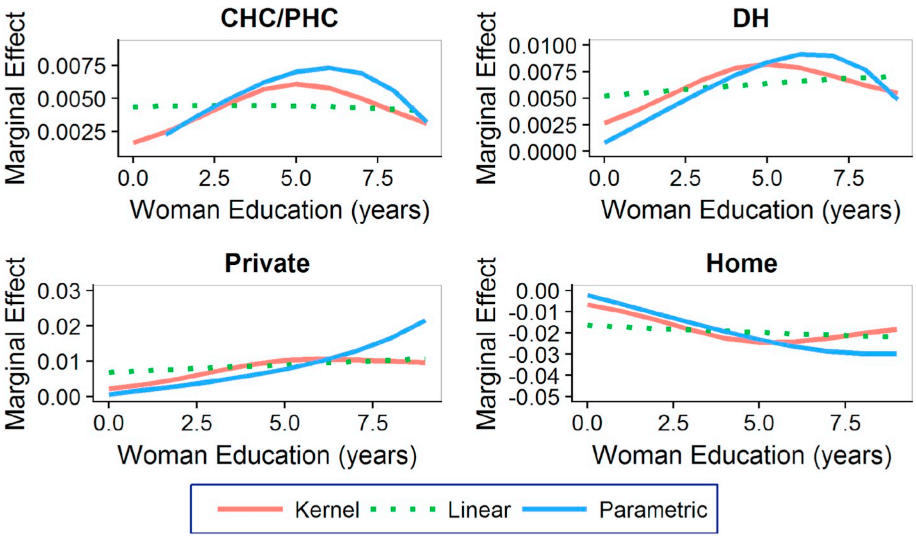


Fig. 7. Mother Education and Distance to DH graphs.

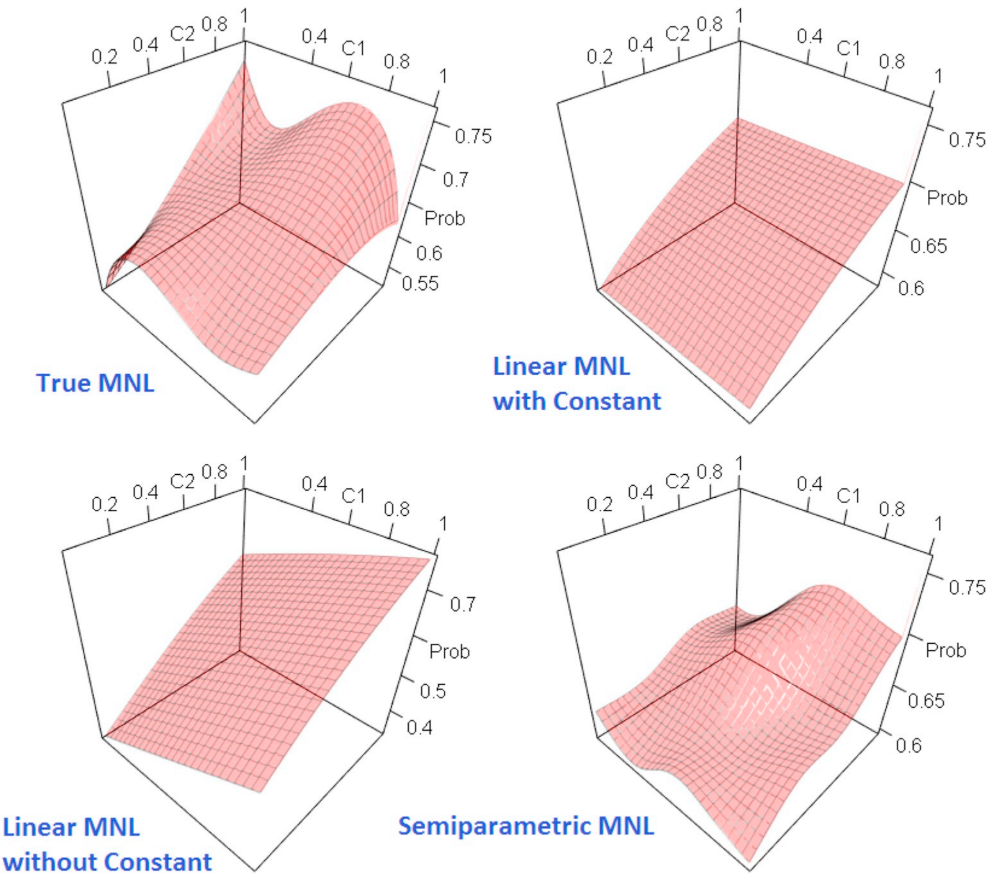


Fig. 8. Asset Index and Birth order graphs.

end), whereas they are relatively similar in the middle part of the distribution. The marginal impact of birth order also shows similar patterns – the linear MNL model marginal effect curve is either completely above (Home) or below (DH, Private and CHC/PHC) the kernel MNL model estimates (see Fig. 8d). These further exemplify the variance in the results from the two models. Despite these differences, the sign of the marginal effects for all four birth location choices is the same across the two specifications when looking at the impact of distance to district hospital and birth order.

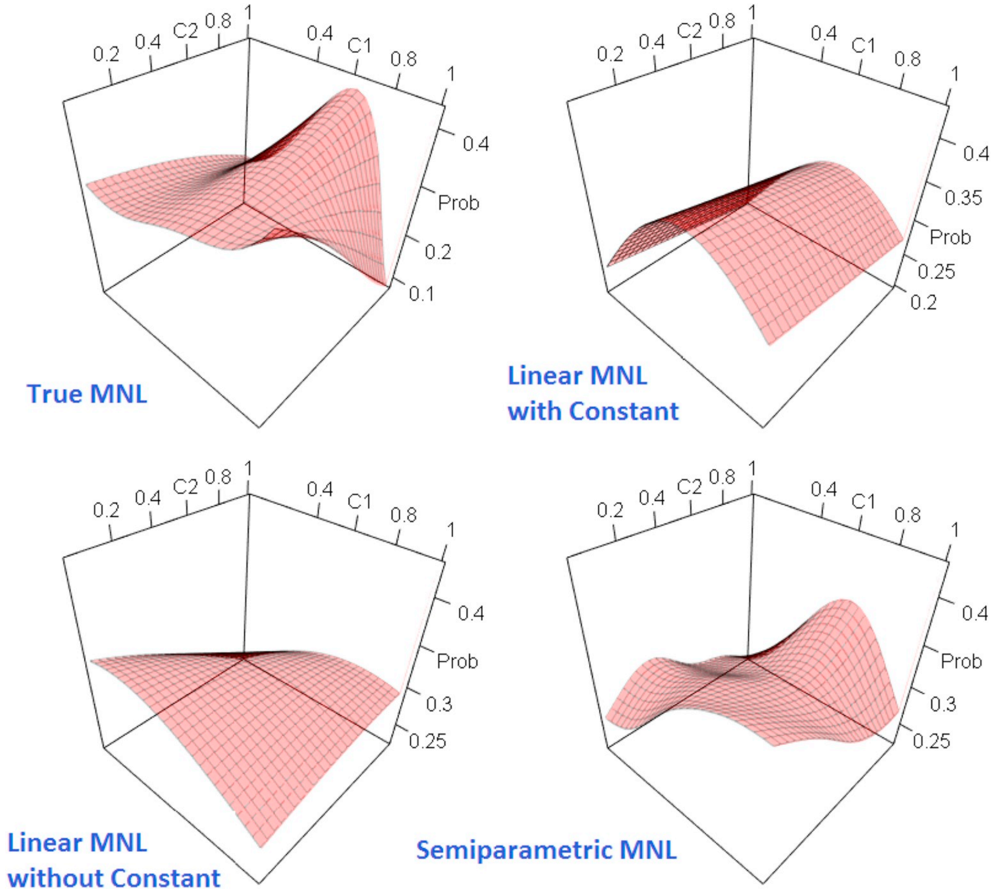


Fig. 9. Parametric Specification mimicking Kernel MNL.

As demonstrated in the discussion above, there are substantial differences in the marginal effect estimates from the linear and kernel MNL frameworks. However, these differences bring a couple of questions – should practitioners stick to the conventional parametric models or switch to more flexible estimation approaches? And then how should practitioners use the results from the kernel-based models to make inferences?

Whereas kernel-based models can uncover the heterogeneous effects of covariates, parametric models are superior in interpretability (due to closed form expressions of marginal effects). We propose the following strategy to leverage the advantages of both models – find the parametric specification that most closely mimics the kernel MNL marginal effects. This strategy would enable ease of inference (parametric model), while allowing for flexibility in the way different factors are modelled (kernel specification).

As an illustration, we perform the aforementioned exercise for the maternal education variable in our setup. We tried various transformations of this variable including polynomials of different orders, as well as exponential and logarithmic transformations. The results are shown in Fig. 9 – the parametric model with quadratic and quartic power polynomial of maternal education closely corresponds to the results from the kernel MNL. This shows that not only can the kernel-based estimates be recovered from some parametric specification, it also illustrates that the functional form is relatively simple. This potentially points toward the use of the kernel results as a benchmark for specification searches/choices among different parametric models.

7. Conclusion

We analyze the role of four critical factors (maternal education, household assets, distance to formal health facility, and birth order) in determining the choice of location for child delivery using a nationally representative revealed-preference dataset from India (N = 157,804). The choice set that we use in our analysis has four options, namely Home Births, District Hospital (DH), Private Clinics, and CHC/PHC. Our results indicate that maternal education and household assets have a positive impact on the probability of institutional birth (non-home birth), whereas, as expected, distance to health facilities and birth order have a negative association. However, due to its restrictive assumptions, the linear MNL model is unable to uncover heterogeneous effects of these variables. To study these non-linear effects, we use a semi-parametric kernel MNL specification and compare it with the linear MNL model.

The kernel specification enables us to flexibly estimate the relationship between covariates and the outcome of interest by allowing a subset of covariates to enter the structural link equation in a non-parametric manner. The estimation of the kernel MNL model follows a profile likelihood approach that involves iterative updating of parameters, and is challenging due to the rapid increase in the number of parameters with a rise in sample size. We thus propose a generalized parallel estimation strategy to handle larger datasets, a problem that is often encountered in health economics (among other disciplines with large datasets).

The use of semiparametric models is also limited due to difficulties in deriving policy insights in a straightforward manner, which sometimes overrides the gain in flexibility offered by these specifications. We overcome this challenge and present a step-by-step procedure to analyze the kernel MNL estimates. We create 3D plots to visualize the interaction and individual effects of two covariates on the probability of choosing a healthcare facility, and then create 2-D plots that help identify the corresponding univariate variation. Subsequently we compute marginal effects by plotting the slope of the 2-D plots against the covariate distribution. We find considerable non-linearity (heterogeneity) in the effects of the aforementioned covariates on choice. Due to the parametric assumptions, the linear MNL masks these patterns.

Further, through a Monte Carlo study we show that the kernel MNL outperforms the linear MNL model based on different statistical criteria, which further raises the strength of our findings. Finally, we reverse-engineer the parametric model that most closely mimics the semi-parametric model results. This helps provide practitioners with an analytical (parametric) specification that plausibly models the data generating process more accurately (like the kernel model). We thus make a strong case for the use of semi-parametric kernel MNL in modeling unordered categorical outcomes.

References

Abe, M., 1999. A generalized additive model for discrete-choice data. J. Bus. Econ. Stat. 17 (3), 271–284.

Alkema, Leontine, Chou, Doris, Hogan, Daniel, Zhang, Sanqian, Moller, Ann-Beth, Gemmill, Alison, Fat, Doris Ma, Boerma, Ties, Temmerman, Marleen, Mathers, Colin, 2016. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group. The Lancet 387 (10017), 462–474. Elsevier. http://www.thelancet.com/pb/assets/raw/Lancet/pdfs/S0140673615008387.pdfLaTEX.

Analytics, R., Weston, S., 2014. doparallel: foreach parallel adaptor for the parallel package. R package version 1 (8).

Aseweh Abor, P., Abekah-Nkrumah, G., Sakyi, K., Adjasi, C.K., Abor, J., 2011. The socio-economic determinants of maternal health care utilization in Ghana. Int. J. Soc. Econ. 38 (7), 628–648.

Baird, S., McIntosh, C., Özler, B., 2011. Cash or condition? evidence from a cash transfer experiment. Q. J. Econ. 126 (4), 1709–1753.

Bartlett III, A., De Bocaletti, M.E.P., Bocaletti, M.A., 1993. Reducing perinatal mortality in developing countries: high risk or improved labour management? Health Policy Plan. 8 (4), 360–368.

Belfield, C.R., Kelly, I.R., 2012. The benefits of breast feeding across the early years of childhood. J. Hum. Cap. 6 (3), 251–277.

Bhatia, J.C., Cleland, J., 1995. Determinants of maternal care in a region of south India. Health Transition Rev. 5 (2), 127–142.

Biks, G.A., Tariku, A., Tessema, G.A., 2015. Effects of antenatal care and institutional delivery on exclusive breastfeeding practice in northwest Ethiopia: a nested case–control study. Int. Breastfeed. J. 10 (1), 30.

Bonu, S., Bhushan, I., Rani, M., Anderson, I., 2009. Incidence and correlates of 'catastrophic'maternal health care expenditure in India. Health Policy Plan. 24 (6), 445–456.

Breierova, L., Duflo, E., 2004. The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less than Mothers? Technical Report. National bureau of economic research.

Campbell, O.M., Calvert, C., Testa, A., Strehlow, M., Benova, L., Keyes, E., Donnay, F., Macleod, D., Gabrysch, S., Rong, L., et al., 2016. The scale, scope, coverage, and capability of childbirth care. Lancet 388 (10056), 2193–2208.

Campbell, O.M., Graham, W.J., steering group, L.M.S.S., et al., 2006. Strategies for reducing maternal mortality: getting on with what works. Lancet 368 (9543), 1284–1299.

Celik, Y., Hotchkiss, D.R., 2000. The socio-economic determinants of maternal health care utilization in Turkey. Soc. Sci. Med. 50 (12), 1797–1806.

Chalumeau, M., Salanave, B., Bouvier-Colle, M.-H., Bernis, L. d., Prua, A., Breart, G., 2000. Risk factors for perinatal mortality in west africa: a population-based study of 20 326 pregnancies. Acta Paediatr. 89 (9), 1115–1121.

Chou, S.-Y., Liu, J.-T., Grossman, M., Joyce, T., 2010. Parental education and child health: evidence from a natural experiment in taiwan. Am. Econ. J. Appl. Econ. 2 (1), 33–61.

Chowdhury, M.E., Ronsmans, C., Killewo, J., Anwar, I., Gausia, K., Das-Gupta, S., Blum, L.S., Dieltiens, G., Marshall, T., Saha, S., et al., 2006. Equity in use of home-based or facility-based skilled obstetric care in rural Bangladesh: an observational study. Lancet 367 (9507), 327–332.

Currie, J., Moretti, E., 2003. Mother's education and the intergenerational transmission of human capital: evidence from college openings. Q. J. Econ. 118 (4), 1495–1532.

Currie, J., Vogl, T., 2013. Early-life health and adult circumstance in developing countries. Annu. Rev. Econ. 5 (1), 1–36.

Darmstadt, G.L., Bhutta, Z.A., Cousens, S., Adam, T., Walker, N., De Bernis, L., Team, L.N.S.S., et al., 2005. Evidence-based, cost-effective interventions: how many newborn babies can we save? Lancet 365 (9463), 977–988.

Duflo, E., Dupas, P., Kremer, M., 2015. Education, hiv, and early fertility: experimental evidence from Kenya. Am. Econ. Rev. 105 (9), 2757–2797.

Elo, I.T., 1992. Utilization of maternal health-care services in Peru: the role of women's education. Health Transition Rev. 2 (1), 49–69.

Falkingham, J., 2003. Inequality and changes in women's use of maternal health-care services in Tajikistan. Stud. Fam. Plann. 34 (1), 32–43.

Feyissa, T.R., Genemo, G.A., 2014. Determinants of institutional delivery among childbearing age women in western Ethiopia, 2013: unmatched case control study. PLoS One 9 (5), e97194.

Filippi, V., Ronsmans, C., Campbell, O.M., Graham, W.J., Mills, A., Borghi, J., Koblinsky, M., Osrin, D., 2006. Maternal health in poor countries: the broader context and a call for action. Lancet 368 (9546), 1535–1541.

Fukuda, D., Yai, T., 2010. Semiparametric specification of the utility function in a travel mode choice model. Transportation 37 (2), 221–238.

Gage, A.J., Guirlène Calixte, M., 2006. Effects of the physical accessibility of maternal health services on their use in rural Haiti. Popul. Stud. 60 (3), 271–288.

Garg, C.C., Karan, A.K., 2008. Reducing out-of-pocket expenditures to reduce poverty: a disaggregated analysis at rural-urban and state level in India. Health Policy Plan. 24 (2), 116–128.

Gething, P.W., Johnson, F.A., Frempong-Ainguah, F., Nyarko, P., Baschieri, A., Aboagye, P., Falkingham, J., Matthews, Z., Atkinson, P.M., 2012. Geographical access to care at birth in Ghana: a barrier to safe motherhood. BMC Public Health 12 (1), 991.

Grépin, K.A., Bharadwaj, P., 2015. Maternal education and child mortality in Zimbabwe. J. Health Econ. 44, 97–117.

Grossman, M., 2006. Education and nonmarket outcomes. In: Handbook of the Economics of Education. 1. pp. 577–633.

Grytten, J., Skau, I., Sørensen, R.J., 2014. Educated mothers, healthy infants. the impact of a school reform on the birth weight of Norwegian infants 1967–2005. Soc. Sci. Med. 105, 84–92.

Guliani, H., Sepehri, A., Serieux, J., 2012. What impact does contact with the prenatal care system have on women's use of facility delivery? evidence from low-income countries. Soc. Sci. Med. 74 (12), 1882–1890.

Güneş, P.M., 2015. The role of maternal education in child health: evidence from a compulsory schooling law. Econ. Educ. Rev. 47, 1–16.

Gwatkin, D.R., Bhuiya, A., Victora, C.G., 2004. Making health systems more equitable. Lancet 364 (9441), 1273–1280.

Hjortsberg, C., 2003. Why do the sick not utilise health care? the case of Zambia. Health Econ. 12 (9), 755–770.
Kalyanaram, G., Little, J.D., 1994. An empirical analysis of latitude of price acceptance in consumer package goods. J. Consum. Res. 21 (3), 408–418.
Kesterton, A.J., Cleland, J., Sloggett, A., Ronsmans, C., 2010. Institutional delivery in rural India: the relative importance of accessibility and economic status. BMC Pregnancy Childbirth 10 (1), 30.
Kneib, T., Baumgartner, B., Steiner, W.J., 2007. Semiparametric multinomial logit models for analysing consumer choice behaviour. Adv. Stat. Anal. 91 (3), 225–244.
Köhler, M., Schindler, A., Sperlich, S., 2014. A review and comparison of bandwidth selection methods for kernel regression. Int. Stat. Rev. 82 (2), 243–274.
Kramer, M.S., Aboud, F., Mironova, E., Vanilovich, I., Platt, R.W., Matush, L., Igumnov, S., Fombonne, E., Bogdanovich, N., Ducruet, T., et al., 2008. Breastfeeding and child cognitive development: new evidence from a large randomized trial. Arch. Gen. Psychiatr. 65 (5), 578–584.
Krishnamurthi, L., Raj, S., 1988. A model of brand choice and purchase quantity price sensitivities. Market. Sci. 7 (1), 1–20.
Kumar, R., Singh, M., Kaur, M., 1997. Impact of health centre availability on utilisation of maternity care and pregnancy outcome in a rural area of Haryana. J. Indian Med. Assoc. 95 (8), 448–450.
Kumar, S., Dansereau, E.A., Murray, C.J., 2014. Does distance matter for institutional delivery in rural India? Appl. Econ. 46 (33), 4091–4103.
Kusiako, T., Ronsmans, C., Van der Paal, L., 2000. Perinatal mortality attributable to complications of childbirth in matlab, Bangladesh. Bull. World Health Organ. 78 (5), 621–627.
Ladusingh, L., Holendro Singh, C., 2007. Rich–poor gap in maternal care: the case of northeast India. Asian Popul. Stud. 3 (1), 79–94.
Lance, L., et al., 2011. Nonproduction benefits of education. In: Handbook of the Economics of Education. 4. pp. 183–282.
Langrock, R., Heidenreich, N.-B., Sperlich, S., 2014. Kernel-based semiparametric multinomial logit modelling of political party preferences. Stat. Methods Appl. 23 (3), 435–449.
McFadden, D., 1974. The measurement of urban travel demand. J. Public Econ. 3 (4), 303–328.
MOHFW, 2016. Household Health Expenditure in India (2013-14). NSHRC,Ministry of Health and Family Welfare, Government of India.
Müller, M., 2001. Estimation and testing in generalized partial linear models–a comparative study. Stat. Comput. 11 (4), 299–309.
Nair, M., Ariana, P., Webster, P., 2012. What influences the decision to undergo institutional delivery by skilled birth attendants? a cohort study in rural Andhra Pradesh, India. Rural Rem. Health 12 (4) 2311–2311.
Navaneetham, K., Dharmalingam, A., 2002. Utilization of maternal health care services in southern India. Soc. Sci. Med. 55 (10), 1849–1869.
Oddy, W.H., Kendall, G.E., Li, J., Jacoby, P., Robinson, M., De Klerk, N.H., Silburn, S.R., Zubrick, S.R., Landau, L.I., Stanley, F.J., 2010. The long-term effects of breastfeeding on child and adolescent mental health: a pregnancy cohort study followed for 14 years. J. Pediatr. 156 (4), 568–574.
Odiit, A., Amuge, B., 2003. Comparison of vaccination status of children born in health units and those born at home. East Afr. Med. J. 80 (1), 3–6.
Organization, W.H., 2006. The World Health Report 2006: Working Together for Health. World Health Organization.
Peters, D.H., 2002. Better Health Systems for India's Poor: Findings, Analysis, and Options. World Bank Publications.
Quigley, M.A., Hockley, C., Carson, C., Kelly, Y., Renfrew, M.J., Sacker, A., 2012. Breastfeeding is associated with improved child cognitive development: a population-based cohort study. J. Pediatr. 160 (1), 25–32.
Raghupathy, S., 1996. Education and the use of maternal health care in Thailand. Soc. Sci. Med. 43 (4), 459–471.
Ronsmans, C., Graham, W.J., steering group, L.M.S.S., et al., 2006. Maternal mortality: who, when, where, and why. Lancet 368 (9542), 1189–1200.
Sarma, S., 2009. Demand for outpatient healthcare. Appl. Health Econ. Health Policy 7 (4), 265–277.
Sawhney, N., 1993. Management of Family Welfare Programme in uttar pradesh: Infrastructure Utilization Quality of Services Supervision and MIS.
Setegn, T., Gerbaba, M., Belachew, T., 2011. Determinants of timely initiation of breastfeeding among mothers in goba woreda, south east Ethiopia: a cross sectional study. BMC Public Health 11 (1), 217.
Severini, T.A., Wong, W.H., 1992. Profile likelihood and conditionally parametric models. Ann. Stat. 20 (4), 1768–1802.
Short, S.E., Zhang, F., 2004. Use of maternal health services in rural China. Popul. Stud. 58 (1), 3–19.
Skordis-Worrall, J., Pace, N., Bapat, U., Das, S., More, N.S., Joshi, W., Pulkki-Brannstrom, A.-M., Osrin, D., 2011. Maternal and neonatal health expenditure in Mumbai slums (India): a cross sectional study. BMC Public Health 11 (1), 150.
Toan, N., Hoa, H., Trong, P., Höjer, B., Persson, L., Sundström, K., 1996. Utilisation of reproductive health services in rural vietnam; are there equal opportunities to plan and protect pregnancies? J. Epidemiol. Community Health 50 (4), 451–455.
UN, 2014. The millennium development goals report.
UNICEF, 2018. Levels & Trends in Child Mortality (by UN interagency group for child mortality estimation). https://childmortality.org/wp-content/uploads/2018/12/UN-IGME-Child-Mortality-Report-2018.pdf.
WHO, 2015. World Health Statistics. World Health Organization, Geneva.