Modeling fertility pattern in Afghanistan using individual, familial, social, and regional factors by Bayesian methods

Hao Xu

August 31, 2024

1 Introduction

Fertility analysis plays a vital role in understanding demographic patterns, health outcomes, and societal progress. By studying fertility, researchers gain valuable insights into population growth, aging, and distribution, which in turn informs policymakers on how to effectively allocate resources, plan for future needs, and address public health concerns.

In this study, we focus on modeling women's fertility patterns in Afghanistan, examining a variety of factors that contribute to fertility outcomes. These factors include individual characteristics such as age and education, familial dynamics like the household head's age, socioeconomic indicators such as wealth and media exposure, and regional differences across provinces and between urban and rural areas. Using a Bayesian hierarchical Poisson regression model, we have reached the conclusion that these factors are significantly associated with women's fertility patterns. However, the Poisson regression model assumes that the mean and variance are equal, which is often not true in real-life data. Fertility rates, often

measured by the number of children ever born (CEB), typically exhibit overdispersion or underdispersion, making the Poisson model inadequate.

To address this limitation, we propose using the Conway-Maxwell-Poisson (COM-Poisson) regression model, which can handle both overdispersion and underdispersion in count data. By applying this model, we aim to achieve a more accurate and nuanced understanding of fertility dynamics in Afghanistan.

1.1 The objective of this study

The core objective of this research is to develop a more robust model that can handle both overdispersion and underdispersion in fertility data. We propose a novel Bayesian hierarchical weighted Conway-Maxwell-Poisson (COM-Poisson) regression model to analyze the number of children ever born (CEB) among married women of reproductive age in Afghanistan. Additionally, we will evaluate and compare the model fit and robustness of the proposed model against the Bayesian hierarchical Poisson model, with particular attention to improving the treatment of underdispersion.

From an applied perspective, this is the first time a Bayesian hierarchical COM-Poisson model has been employed in fertility analysis. This approach allows us to model dispersed count data more accurately, providing a better fit for real-world fertility data in Afghanistan.

Methodologically, this study breaks new ground by incorporating Bayesian hierarchical modeling techniques with the COM-Poisson regression—an innovative combination not previously explored in fertility research. In addition to implementing this model, we will compare its performance with that of the Bayesian hierarchical Poisson regression model, focusing on differences in model fit and robustness agains prior miss-specification.

Afghanistan has endured decades of conflict, significantly impacting its demographic and social landscape. Time permitting, we will also explore fertility patterns prior to the onset of

war and compare them with recent post-war statistics, providing insights into how prolonged conflict has influenced fertility trends.

In summary, the study aims to make contributions across three key areas:

1. Develop and implement the Bayesian hierarchical weighted COM-Poisson model and apply it to fertility analysis in Afghanistan.

2. Evaluate the effects of individual, household, and regional factors on the number of children ever born (CEB) using the proposed model and compare the findings with those of previous models. Regional differences are captured through hierarchical modeling.

3. Assess model fit and robustness, using the Bayesian Information Criterion (BIC) for model evaluation, and testing robustness by altering prior distributions. Comparisons of model fit and robustness will be made with existing models.

By analyzing how factors like education, economic status, geographic location, and familial status influence fertility, the study will provide valuable insights that can guide the design of targeted interventions. These interventions can help manage population growth and achieve broader development goals, while also ensuring that maternal and child health programs, family planning, and reproductive health education are tailored to the specific needs of different communities in Afghanistan.

The study's focus on regional differences in fertility patterns is particularly important for addressing disparities accross geographic areas. By identifying how factors such as women's age, education, and media access vary across regions, the study can inform more equitable distribution of resources and support. This helps ensure that underserved areas receive the attention they need, leading to more balanced development and improved health outcomes across the country.

From a theoretical standpoint, this research fills a gap in the existing literature by employing

a novel hierarchical COM-Poisson model and applying to recent fertility data. Its ability to handle both overdispersed and underdispersed count data will provide new insights into fertility trends in conflict-affected and developing regions such as Afghanistan.

2 Data Description

In this reaseach, we utilize data from the Multiple Indicator Cluster Surveys (MICS) Round 6 Datasets in Afghanistan. Data of Individual woman and their household information were collected over the span of 2022 and 2023 by the United Nations International Children's Emergency Fund (UNICEF) along with the National Statistics and Information Authority (NSIA). The survey took sampling frame based on 2019 satellite imagery and includes urban and rural areas across 34 provinces, with the objective of obtaining reliable national, urban, rural, and provincial estimates. Among these, 32 provinces with both urban and rural areas along with 2 rural provinces were used as strata. In total, the dataset was grouped into 34 rural strata and 32 urban strata. The survey was a cross-sectional survey with two-staged stratified clustering. On the first stage, Primary Sampling Units (PSUs) were selected, which is also called enumeration areas (EAs). Five provinces — namely Kabul, Nangarhar, Balkh, Kandahar and, Herat — were identified as being larger than the others and were given larger sample populations. 28 enumerated areas(EA) were given if it was a small province, 34 if it was large. Within each EA, a random sample of 24 households were selected for sampling. Data is weighted to account for the probabilities of selection at each stage, ensuring that results can be generalized to the entire population or specific subgroups. Probability Sampling: Each household and member had a known probability of selection. This allows for valid inferences about the population by weighting data according to the inverse of selection probabilities. Further information of the data can be found in the survey (UNICEF, 2023).

2.1 Data Selection and Merging

The UNICEF MICS dataset includes 44,874 Afghan women aged 15 to 49. Of these, 16,309 women had no recorded data on the number of children ever born (CEB). Specifically, 15,776 women had never married, and 533 women had missing CEB data along with other key predictors, making their records unusable. For this study, the fertility sample was adjusted to 28,563 women by excluding the 15,776 unmarried women and the 533 women with incomplete data. In Afghan society, childbearing is primarily expected to occur within marriage due to strong cultural, religious, and social norms. Therefore, excluding unmarried women is appropriate for fertility studies, as their childbearing patterns do not reflect societal expectations. Additionally, two women with missing data on their age at first marriage were also excluded to ensure the dataset's completeness for the model.

In this study, the number of children ever born per woman (CEB) is taken as the response variable. The explanatory variables, as shown in Table 1, were selected based on prior research indicating their potential impact on the number of children ever born (CEB), as well as variables of interest for deeper exploration. These variables fall into four broad categories:

1. Individual factors: These include the woman's age (in years), her age at first marriage (in years), and her education level (whether she has received schooling).

2. Familial factors: These cover the household head's age (in years), the household head's education level (whether they have received schooling), the household head's sex (male or female), and polygyny (whether the husband has other wives).

3. Social factors: These consist of the wealth index (categorized into poor, middle, and rich based on economic quintiles) and media access (whether the woman has any exposure to media).

4. Regional factors: These include the province (with 34 provinces in Afghanistan) and whether the area is urban or rural.

Havig boardly categorized predictors allows for a more organized analysis of how individual, familial, social, and regional factors affect fertility patterns in Afghanistan.

Regarding the woman's age, previous research has noted a quadratic relationship between a woman's age and the number of children ever born (Tomal et al., 2022). Also, a linear term would assume a constant rate of change across all ages, which doesn't capture this more complex pattern. By including a quadratic term, the model can account for this non-linear relationship, capturing both the rise and the decline in fertility as women age. Therefore, we will also include a quadratic term for age in our regression model. This means that the number of children ever born typically increases with age up to a certain point, and then decreases as women move beyond their peak childbearing years. This allows for a more accurate and realistic representation of fertility patterns, aligning with prior findings from Tomal et al. (2022).

The categorical variables listed in Table 1 were not directly taken from the dataset but were subject to preprocessing. Due to the hierarchical modeling approach and the fact that not all categories are represented in each stratum, some categories had to be merged. The preprocessing steps were as follows: Media exposure was originally measured on a scale of 0 to 3 for television and radio usage frequency (0 meaning never, 3 indicating daily use). These values were summed to create a 0-9 scale reflecting the woman's interaction with mass media. Since values above 3 were rare, the data was converted into a binary variable, where 1 indicated any media exposure and 0 indicated none. Missing data for the number of other wives a woman's husband had was imputed as 0, as this question followed a query about whether the husband had any other wives. For education, we used a follow-up question on whether the woman had ever attended school, replacing missing values with 0 under the assumption that the woman had no schooling. The responses were initially recorded as a

scale from 1 (primary attendance) to 5 (Islamic schooling), but due to the sparse data on schooling beyond the primary level, the data was simplified to a binary variable: 0 for no schooling and 1 for having attended school. Islamic education was categorized as 0 for the purposes of this analysis.

Table 1: Descriptive statistics for the predictors used in this study.

Variables	Type	Mean	SD
R—Women's age	Quantitative	31.2	8.6
R—Women's age squared	Quantitative	1048.4	565.3
R—Household head's age	Quantitative	44.3	14.9
R—Household head's age squared	Quantitative	2183.6	1466.1
R—Age at marriage	Quantitative	18.3	4.3
Variables	**Categories**	**Frequency**	**Percentage**
F—Province (33 predictors)	Herat as reference, See Fig. 1 for frequency of each province		
F—Area	Urban	4722	16.5
	Rural (reference level)	23842	83.5
R—Education of Woman	No(reference level)	23705	83.0
	Yes	4858	17.0
R—Household head's Edu level	No (reference level)	18789	65.8
	Yes	9774	34.2
F—Household head's sex	Male (reference level)	18789	96.1
	Female	1127	3.9
F—Number of other wives	0 (reference level)	26074	91.3
	≥ 1	2489	8.7
R—Wealth index	Poor (category 1)	12562	44.0
	Middle (reference level)	6388	22.3
	Richest (category 3)	9613	33.7
R—Media access	No Access (reference level)	19215	67.3
	Some Access	9348	34.3

The First Capital letter in the Variables name indicate the type of the predictor: Fixed(F) or Random(R)

While performing the hierarchical regression, it was discovered that many of the strata contained only a single value for key variables such as media access and polygyny. This indicates a very strong homogeneity within these strata, creating a challenge since the random variables do not exhibit true randomness. To address this issue, we treated variables with only one value across the whole stratum as fixed predictors instead of random predictors.

The reference level for the area was set to be rural, as two provinces (Panjsher and Nooristan) had no urban population. The selection of reference level aligns our analytical framework with the general demographic distribution and ensures a consistent basis for comparison across different regions and living conditions.

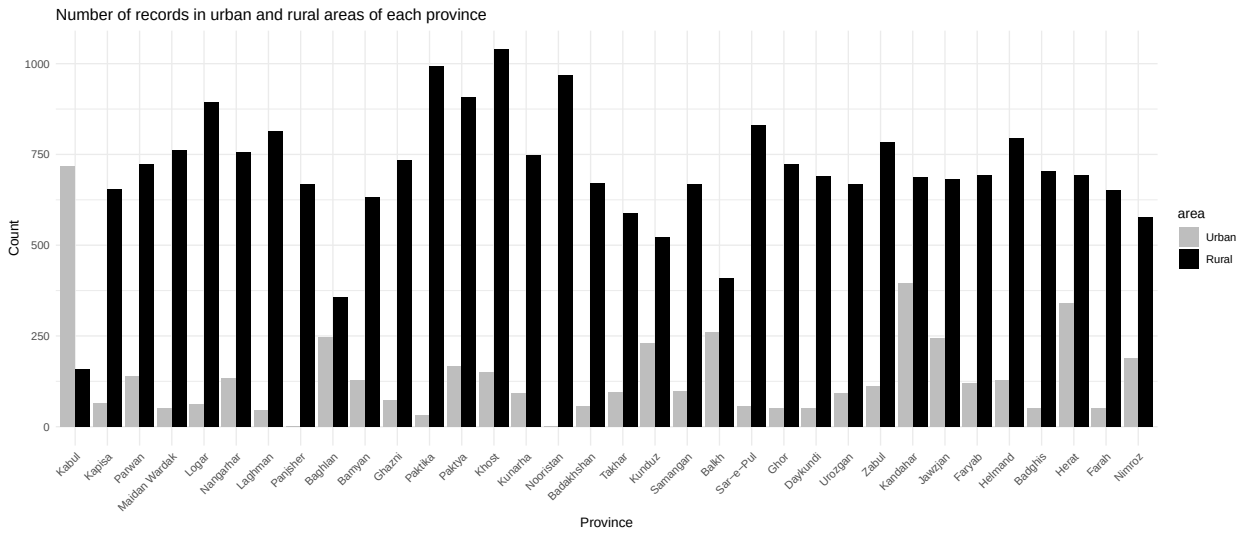


Figure 1: Number of women in each stratum (Province and Area)

3 Literature Review

3.1 Development of Probability Models

Traditional approaches to modeling count data, such as the number of children ever born (CEB), often begin with Poisson regression. This model is a type of generalized linear model (GLM) that uses the exponential function as its link function. For example, a study on fertility determinants in Nigeria successfully employed a Poisson regression model to predict CEB based on women's background characteristics, identifying key factors such as age, education, and age at first marriage (Fagbamigbe and Adebowale, 2014). While classic Poisson regression is straightforward and effective for handling count data, it falls short when dealing with discrepancies across regional characteristics, particularly in hierarchical or nested data structures.

To address these limitations, more advanced methods like linear mixed-effects models (also known as hierarchical models) have been developed. These models extend traditional GLMs by incorporating random effects, allowing for the modeling of variability at different levels, such as within regions or provinces. Mixed-effects models offer significant advantages over

repeated measures ANOVAs and multiple regression, as they account for dependencies in the data, handle missing data more effectively, and provide more accurate estimates of the effects of interest (Brown, 2021). This makes them particularly valuable in demographic studies, where data are often complex and hierarchical.

In recent years, Bayesian methods have gained attention for their ability to further enhance the analysis of hierarchical models, particularly in fields that involve complex data structures, such as demographics and social sciences. Hierarchical modeling allows for the appropriate consideration of intra-stratum correlations and the heterogeneous influence of global-level units on individual outcomes (Erkan Ozkaya et al., 2013). Bayesian methods have further enhanced these models by incorporating prior knowledge and providing greater flexibility in handling unbalanced data, which is often encountered in real-world datasets (Britten et al., 2021).

In settings like Afghanistan, where data are often sparse, noisy, or incomplete, Bayesian hierarchical models can offer distinct advantages by accommodating expert insights and integrating uncertainty through prior distributions Specifically, Bayesian hierarchical models are well-suited to capture the complex interactions between individual, familial, and regional factors, offering a more nuanced understanding of fertility dynamics in diverse and challenging environments. This study will explore the use of Bayesian methods alongside traditional approaches, highlighting their potential to improve the accuracy and reliability of fertility predictions in such contexts.

3.2 Determinants of Fertility Pattern in Afghanistan

The number of children ever born (CEB) in a family is influenced by many variables such as the education of both the mother and father, the country's economy, religious beliefs, and access to contraceptive methods. Previous studies (Osmani et al., 2015; Tomal et al., 2022) have emphasized the complex interactions between these factors, and highlight the need for

data driven, targeted policies and programs that can address the influences of these variables. Throughout Afghanistan's history, young girls have been particularly vulnerable to early marriage practices, which not only violate their human rights but also have negative impacts on their mental well-being. The level of education plays a crucial role in determining the prevalence of child marriage. A recent study based on the 2015 Afghanistan Demographic and Health Survey (Ehsan et al., 2021) examined the influence of poverty and education on early marriages among girls. Through the use of binary logistic regression analysis, the research revealed that girls from impoverished households were 1.18 times more likely to experience early marriage. Similarly, girls who did not attend school had a 1.16 times higher likelihood of getting married between the ages of 10 to 19. Other factors such as illiteracy, living in rural areas, belonging to specific ethnic groups, and regional disparities also contribute significantly to the high rates of child marriages in Afghanistan. The study from Dadras et al. (2022) investigates the impact of child marriage on reproductive outcomes and healthcare utilization among young Afghan women, using data from the 2015 Afghanistan Demographic and Health Survey. It focuses on women aged 20-24, revealing that 52% of them were married before 18. Poverty and illiteracy are key factors contributing to early marriage. The study finds that child marriage is linked to negative reproductive outcomes, such as rapid repeat childbirth, lack of skilled birth attendance, and fewer institutional deliveries. Analysis of the factors affecting fertility rates and access to reproductive health can develop policies to assist with several goals for the countries development.

Studies on the factors affecting fertility in Afghanistan do exist, though they are very few and often focus on single factors, such as the impact of domestic violence (Shams Ghahfarokhi, 2024), or rely on older datasets like the 2010 Afghanistan Mortality Survey (AMS) (Indian Institute for Health Management Research (IIHMR), 2013) and the 2015 Afghanistan Demographic and Health Survey (ADHS) (Central Statistics Organization (CSO), 2017). No research has yet been conducted using the UNICEF MICS dataset, which our study will utilize. To explore potential avenues for analysis, we can also look to research in countries

with similar cultural or geographic contexts. Fertility rates in South Asia, particularly in Bangladesh, have been extensively researched (Tomal et al., 2022), revealing the significant influence of factors such as education, family planning access, and socioeconomic status. This established framework offers valuable insights for studying fertility in other countries. Similarly, studies in neighboring countries of Afghanistan, including Iran and Pakistan, highlight comparable trends. In Iran, fertility rates have shown a significant decline over the years, influenced by increased education levels and widespread family planning programs (Abbasi-Shavazi et al., 2015; Hosseini-Chavoshi et al., 2017). Pakistan also exhibits similar patterns, with research indicating the crucial role of socioeconomic factors and access to reproductive health services in shaping fertility rates (Aslam et al., 2016). Understanding these regional and cultural dynamics and their impacts on fertility is essential for developing targeted policies that address the unique challenges and opportunities in improving reproductive health outcomes across South Asia.

3.3 The Conway-Maxwell-Poisson Model

The Conway-Maxwell-Poisson (CMP or COM-Poisson) distribution is a two-parameter generalisation of the Poisson distribution, allowing for over-dispersion or under-dispersion(Conway and Maxwell, 1962).

For a random variable $Y \sim \text{COM-Poisson}(\lambda, \nu)$, the probability mass function is give by:

$$P(Y = y|\lambda, \nu) = \frac{1}{Z(\lambda, \nu)} \frac{\lambda^y}{(y!)^\nu}, \qquad y \in \mathbb{N} = \{0, 1, 2, \ldots\}, \tag{1}$$

where $Z(\lambda, \nu)$ is a normalizing constant defined by

$$Z(\lambda, \nu) = \sum_{i=0}^{\infty} \frac{\lambda^i}{(i!)^\nu}.$$

The ratio of probabilities of successive integers is

$$\frac{P\{Y = y-1|\lambda,\nu\}}{P\{Y = y|\lambda,\nu\}} = \frac{y^\nu}{\lambda} \tag{2}$$

When $\nu = 1$, the COM-Poisson distribution is degenerate to a Poisson distribution. Values of $\nu > 1$ indicate under-dispersion relative to the Poisson, while $\nu < 1$ indicates over-dispersion. As $\nu \to \infty$, $Z(\lambda,\nu) \to 1+\lambda$, and the COM-Poisson approaches a Bernoulli distribution with a Bernoulli distribution with $P\{X = 1|\lambda,\nu\} = \frac{\lambda}{\lambda+1}$.

When $\nu = 0$ and $\lambda < 1$, $Z(\lambda,\nu)$ is a geometric sum:

$$Z(\lambda,\nu) = \sum_{j=0}^{\infty} \lambda^j = \frac{1}{1-\lambda}, \tag{3}$$

and the distribution itself is geometric:

$$P\{Y = y|\lambda,\nu\} = \lambda^y(1-\lambda), \text{for } y = 0, 1, 2, \ldots \tag{4}$$

When $\nu = 0$ and $\lambda \geq 1, Z(\lambda,\nu)$ does not converge, and the distribution is undefined.

4 Methodology: Bayesian Hierarchical Weighted Poisson Regression

In this section, we present a detailed description of the Bayesian hierarchical weighted Poisson regression model, which incorporates both fixed and random effects from the predictors. This approach leverages the hierarchical nature of the data to provide a more nuanced and comprehensive analysis. Building upon this foundation, we develop a Bayesian hierarchical

model to further explore the structure of the data. We then discuss the preliminary results and address the limitations of the current Poisson assumptions, which lead to the proposal of a new model. The details of this new model, to be integrated into the Bayesian framework, will be elaborated on in the next section.

4.1 The Sampling Distribution

Let m counts the number of strata in data which is indexed by j.

We further consider that there are n_j records of married women in the j^{th} stratum, where each record is indexed by i. Let Y_{ij} be the response variable which counts the number of children ever born for the i^{th} woman in the j^{th} stratum. We consider that Y_{ij} follows a Poisson distribution with mean $\lambda_{ij} > 0$ so that the probability mass function is

$$P(Y_{ij} = y_{ij}|\lambda_{ij}) = \frac{e^{-\lambda_{ij}}{\lambda_{ij}}^{y_{ij}}}{y_{ij}!}, \tag{5}$$

for $y_{ij} \in \{0, 1, 2, 3, ...\}$. The value of λ_{ij} is not constant but varies systematically due to factors like individual characteristics, familial dynamics, socioeconomic status and regional differences.

To fit a Poisson regression model, we use the log-link function, which means that the log of the average birth rate is a linear combination of the fixed and random predictors. The model we consider in this paper is

where

- λ_{ij} is the expectation of Y_{ij}.
- $\mathbf{z}_{ij} = (1, z_{ij}^{(2)}, \ldots, z_{ij}^{(q)})$ is a $(1 \times q)$ is a vector of *random* predictors for which the first element is one ($z_{ij}^{(1)} = 1$).

- $\mathbf{x}_{ij} = (1, x_{ij}^{(2)}, \ldots, x_{ij}^{(q)}, x_{ij}^{(q+1)}, \ldots, x_{ij}^{(p)})$ is a $(1 \times p)$ is a vector of *fixed* predictors concatenated with random predictors for which the first element is one ($x_{ij}^{(1)} = 1$). Furthermore, $x_{ij}^{(t)} = z_{ij}^{(t)}, \forall t \in \{1, 2, \ldots, q\}$.

- $\boldsymbol{\beta}_j = (\beta_{1j}, \beta_{2j}, \ldots, \beta_{qj})^{\mathrm{T}}$ is a $(q \times 1)$ vector of random coefficients for the j^{th} stratum with $E(\boldsymbol{\beta}_j) = \boldsymbol{\theta}_1$. Given $\boldsymbol{\theta}_1$ and Σ, we consider

  $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m \sim$ i.i.d multivariate normal$(\boldsymbol{\theta}_1, \Sigma)$.

- $\boldsymbol{\gamma}_j = (\gamma_{1j}, \gamma_{2j}, \ldots, \gamma_{qj})^{\mathrm{T}}$ is a $(q \times 1)$ is a vector of random effects for the random predictors specific to the j^{th} stratum such that $\boldsymbol{\gamma}_j = \boldsymbol{\beta}_j - \boldsymbol{\theta}_1$. Therefore,

  $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m \sim$ i.i.d multivariate normal$(\mathbf{0}, \Sigma)$.

- $\boldsymbol{\theta}_1 = (\theta_{11}, \ldots, \theta_{1q})^{\mathrm{T}}$ is the q-vector of fixed effects from the random predictors.

- $\boldsymbol{\theta}_2 = (\theta_{21}, \ldots, \theta_{2(p-q)})^{\mathrm{T}}$ is the $(p-q)$-vector of fixed effects from the fixed predictors.

- $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^T = (\theta_{11}, \ldots, \theta_{1q}, \theta_{21}, \ldots, \theta_{2(p-q)})^{\mathrm{T}}$ is the p-vector of fixed effects from the random and fixed predictors placed together. For simplicity, we write $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$.

In addition, we use the following notations:

- $\mathbf{Y}_j$ represents the vector of response variable values in the j^{th} stratum, such that $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, \ldots, Y_{n_j j})^{\mathrm{T}}$.

- $\mathbf{y}_j$ represents the vector of observed values of the response variable in the j^{th} stratum, such that $\mathbf{y}_j = (y_{1j}, y_{2j}, \ldots, y_{n_j j})^{\mathrm{T}}$.

- $\mathbf{Z}_j$ and $\mathbf{X}_j$ represent the design matrix of random and fixed predictors in j^{th} stratum of order $n_j \times q$ and $n_j \times p$, respectively. We sequentially stack $\mathbf{z}_{ij}$ and $\mathbf{x}_{ij}$ over rows, $i = 1, 2, \ldots, n_j$ to get $\mathbf{Z}_j$ and $\mathbf{X}_j$, respectively.

- $\mathbf{Z}$ and $\mathbf{X}$ represent the design matrix of random and fixed predictors putting all the strata together respectively. We get this by sequentially stacking $\mathbf{z}_{ij}$ and $\mathbf{x}_{ij}$ over rows, $i = 1, 2, \ldots, n_j$ and $j = 1, 2, \ldots, m$.

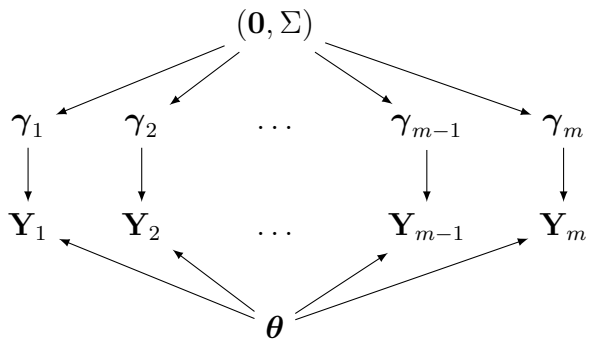


Figure 2: Schematic for the Poisson mixed-effects hierarchical model.

Considering these modeling aspects, the weighted sampling distribution of $\mathbf{Y}_j$ within the jth stratum is written as

$$\begin{aligned} & P(\mathbf{Y}_j = \mathbf{y}_j | \mathbf{Z}_j, \mathbf{X}_j, \boldsymbol{\gamma}_j, \boldsymbol{\theta}) \\ =& \prod_{i=1}^{n_j} \text{dpois}(Y_{ij} = y_{ij} | \lambda_{ij}))^{w_{ij}} \\ =& \prod_{i=1}^{n_j} \text{dpois}(Y_{ij} = y_{ij} | \exp\left(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta}\right))^{w_{ij}} . \end{aligned} \tag{6}$$

The weighted sampling distribution of $\mathbf{Y}$ is:

$$\begin{aligned} & P(\mathbf{Y} = \mathbf{y} | \mathbf{Z}, \mathbf{X}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}) \\ =& \prod_{j=1}^{m} P(\mathbf{Y}_j = \mathbf{y}_j | \mathbf{Z}_j, \mathbf{X}_j, \boldsymbol{\gamma}_j, \boldsymbol{\theta}) \\ =& \prod_{j=1}^{m} \prod_{i=1}^{n_j} \text{dpois}(Y_{ij} = y_{ij} | \exp\left(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta}\right))^{w_{ij}} . \end{aligned} \tag{7}$$

We will consider $\boldsymbol{\gamma}_j$ to be from a multivariate normal distribution. To simplify the notations, if a random vector $\mathbf{Y}$ follows a multivariate normal with mean of $\boldsymbol{\mu}$ and covariance matrix Σ, we write the probability density of $P(\mathbf{Y} = \mathbf{y}|\boldsymbol{\mu}, \Sigma) = \text{dmvnorm}(\mathbf{y}, \boldsymbol{\mu}, \Sigma)$. The sampling distribution of $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m$ is

$$\begin{aligned} & P(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m|\mathbf{0}, \Sigma) \\ =& \prod_{j=1}^{m} \text{dmvnorm}(\boldsymbol{\gamma}_j, \mathbf{0}, \Sigma) \\ =& \prod_{j=1}^{m} (2\pi)^{-q/2}|\Sigma|^{-1/2} \exp\{-\frac{1}{2}\boldsymbol{\gamma}_j^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\gamma}_j\}. \end{aligned} \tag{8}$$

Combining Eq. 8 and Eq. 7, the sampling distribution for the data becomes

$$\begin{aligned} & P(\mathbf{Y} = \mathbf{y}, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_m|\mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}, \Sigma) \\ =& P(\mathbf{Y} = \mathbf{y}|\mathbf{Z}, \mathbf{X}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}) \cdot P(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m|\mathbf{0}, \Sigma) \\ =& \prod_{j=1}^{m} \Big[\text{dmvnorm}(\boldsymbol{\gamma}_j, \mathbf{0}, \Sigma) \times \prod_{i=1}^{n_j} \text{dpois}(Y_{ij} = y_{ij}| \exp\left(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta}\right))^{w_{ij}} \Big]. \end{aligned} \tag{9}$$

4.2 Prior Distributions

Before proceeding to a Bayesian analysis to get the posterior distribution $P(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}, \Sigma|\mathbf{y}, \mathbf{Z}, \mathbf{X})$, we first give the prior distributions for Σ and $\boldsymbol{\theta}$.

Prior Distribution of Σ In our model, $\boldsymbol{\gamma}_j \sim$ i.i.d. multivariate normal$(\mathbf{0}, \Sigma)$. Σ is the variance-covariance matrix of $\boldsymbol{\gamma}_j$, which is the random effect of random predictors from data in the j^{th} stratum. We want to choose a semi-conjugate prior distribution for Σ to simplify the posterior distribution. Wishart distribution is a semi-conjugate prior distribution for the precision matrix Σ^{-1}, and so the inverse-Wishart distribution is our semi-conjugate prior

distribution for the variance-covariance matrix Σ.

$$\Sigma \sim \text{inverse-Wishart}(\eta_0, \mathbf{S}_0^{-1}).$$

We take the prior scale matrix $\mathbf{S}_0$ to be equal to the variance-covariance matrix of the first q coefficients of classical Poisson regression fits, but we'll take the prior degrees of freedom η_0 to be $q + 2$, so that the prior distribution of Σ is reasonably diffuse but has an expectation equal to the sample covariance of the classic Poisson regression estimates.

Prior Distribution of $\boldsymbol{\theta}$ For the fixed effects coefficients $\boldsymbol{\theta}$, we consider a multivariate normal prior distribution with mean vector $\boldsymbol{\mu}_0$ (with a dimension of $(p \times 1)$) and covariance matrix Λ_0. We obtain μ_0 using maximum likelihood estimates from a Poisson regression model $Y_{ij} \sim \text{Poisson}(\exp(\mathbf{x_{ij}}\boldsymbol{\theta}))$ We choose Λ_0 to be the inverse of the Fisher information matrix $I_{\mathbf{Y}}^{-1}(\boldsymbol{\theta}) = (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}$ where $\mathbf{W}$ is defined as:

$$\mathbf{W} = \mathbf{W}(\boldsymbol{\theta}) := \text{diag}\left(\exp(\mathbf{x}_{11}\boldsymbol{\theta}), \ldots, \exp(\mathbf{x}_{n_1 1}\boldsymbol{\theta}), \ldots, \exp(\mathbf{x}_{1,m}\boldsymbol{\theta}), \ldots, \exp(\mathbf{x}_{n_m m}\boldsymbol{\theta})\right) \quad (10)$$

The prior distributions are summarized as follows:

$$\Sigma \sim \text{inverse-Wishart}(\eta_0, \mathbf{S}_0^{-1}). \quad (11)$$

$$\boldsymbol{\theta} \sim \text{MVN}(\boldsymbol{\mu}_0, \Lambda_0) \quad (12)$$

4.3 Full Conditional Distribution

The joint posterior distribution of $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}, \Sigma)|$ can be obtained as:

$$\begin{aligned}
&P(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}, \Sigma | \mathbf{y}, \mathbf{Z}, \mathbf{X}) \\
\propto &P(\mathbf{Y} = \mathbf{y}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m | \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}, \Sigma) \times P(\boldsymbol{\theta}) \times P(\Sigma) \\
= &P(\mathbf{Y} = \mathbf{y} | \mathbf{Z}, \mathbf{X}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}) \times P(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m | \mathbf{0}, \Sigma) \times P(\boldsymbol{\theta}) \times P(\Sigma)
\end{aligned} \tag{13}$$

Full conditional distribution of $\boldsymbol{\theta}$ The full conditional distribution of $\boldsymbol{\theta}$ can be derived from Eq. 13 as:

$$\begin{aligned}
&P(\boldsymbol{\theta} | \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \Sigma, \mathbf{y}, \mathbf{Z}, \mathbf{X}) \\
\propto &P(\mathbf{Y} = \mathbf{y} | \mathbf{Z}, \mathbf{X}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}) \times P(\boldsymbol{\theta}) \\
= &\prod_{j=1}^{m} \prod_{i=1}^{n_j} \text{dpois}(Y_{ij} = y_{ij} | \exp\left(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta}\right))^{w_{ij}} \times \text{dmvnorm}(\boldsymbol{\theta}, \boldsymbol{\mu}_0, \Lambda_0) \\
= &\prod_{j=1}^{m} \prod_{i=1}^{n_j} \left(\frac{e^{-\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})} [\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})]^{y_{ij}}}{y_{ij}!} \right)^{w_{ij}} \\
&\times (2\pi)^{-p/2} |\Lambda_0|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_0)^{\mathrm{T}} \Lambda_0 (\boldsymbol{\theta} - \boldsymbol{\mu}_0)\}
\end{aligned}$$

Because of the Poisson density function here, we don't have a standard form of full conditional distribution, we need to use Metropolis algorithm to generate samples to simulate the posterior distribution of $\boldsymbol{\theta}$.

Full conditional distribution of $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m$ The full conditional distribution of $\boldsymbol{\gamma}_j, j = 1, \ldots, m$ can be obtained from Eq. 13 as:

$$
\begin{aligned}
& P(\boldsymbol{\gamma}_j | \boldsymbol{\theta}, \Sigma, \mathbf{y}_j, \mathbf{Z}_j, \mathbf{X}_j) \\
\propto & P(\mathbf{Y}_j = \mathbf{y}_j | \mathbf{Z}_j, \mathbf{X}_j, \boldsymbol{\gamma}_j, \boldsymbol{\theta}) \times P(\boldsymbol{\gamma}_j | \mathbf{0}, \Sigma) \\
= & \prod_{i=1}^{n_j} \text{dpois}(Y_{ij} = y_{ij} | \exp\left(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta}\right))^{w_{ij}} \times \text{dmvnorm}(\boldsymbol{\gamma}_j, \mathbf{0}, \Sigma) \\
= & \prod_{i=1}^{n_j} \left(\frac{e^{-\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})} [\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})]^{y_{ij}}}{y_{ij}!} \right)^{w_{ij}} \times (2\pi)^{-q/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2}\boldsymbol{\gamma}_j^{\mathrm{T}} \Sigma^{-1} \boldsymbol{\gamma}_j\}
\end{aligned}
$$

This expression does not have an standard form, thus it is not possible to use a Gibbs sampler or Monte Carlo method to generate samples. Instead, we consider the Metropolis Algorithm.

Full conditional distribution of Σ Given the prior distribution for Σ (Eq.11), and the likelihood of observing the random effects $\boldsymbol{\gamma}_j$ for each stratum, the posterior distribution for Σ is derived by combining the prior with the likelihood, resulting in another Inverse-Wishart distribution with updated parameters. This is shown as follows:

$$
\begin{aligned}
& P(\Sigma | \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}, \mathbf{y}, \mathbf{Z}, \mathbf{X}) \\
\propto & P(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m | \mathbf{0}, \Sigma) \times P(\Sigma) \\
\propto & \prod_{j=1}^{m} |\Sigma|^{-1/2} \exp(-\frac{1}{2}\boldsymbol{\gamma}_j^{\mathrm{T}} \Sigma^{-1} \boldsymbol{\gamma}_j) \times |\Sigma|^{-(\eta_0+q+1)/2} \exp\left(-\text{tr}(\mathbf{S}_0 \Sigma^{-1})/2\right) \\
\propto & |\Sigma|^{-(\eta_0+m+q+1)/2} \exp\left(-\text{tr}((\mathbf{S}_0 + \mathbf{S}_{\boldsymbol{\gamma}})\Sigma^{-1})/2\right)
\end{aligned}
$$

where $S_{\boldsymbol{\gamma}} = \sum_{j=1}^{m} (\boldsymbol{\gamma}_j)(\boldsymbol{\gamma}_j)^{\mathrm{T}}$ is the sum of squares of the deviations of $\boldsymbol{\gamma}_j$ from $\mathbf{0}$. Incorporating the likelihood into our prior results in the posterior distribution for Σ:

$$\begin{aligned}&\{\Sigma|\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}, \mathbf{y}, \mathbf{Z}, \mathbf{X}\}\\ \sim&\text{Inverse-Wishart}\left(\eta_0 + m, [\mathbf{S}_0 + \mathbf{S}_{\boldsymbol{\gamma}}]^{-1}\right).\end{aligned} \tag{14}$$

We can then use this posterior distribution of Σ and apply Gibbs sampling algorithm to update Σ in each iteration.

4.4 Metropolis Algorithm

Since there is no standard form for the posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}_j$, we use Metropolis Algorithm to generate a sequence of parameter values, whose empirical distribution approximates the posterior distribution.

The goal is to iteratively generate a collection of parameters $\{\boldsymbol{\theta}^{(0)}, \ldots, \boldsymbol{\theta}^{(S)}\}$ and $\{\boldsymbol{\gamma}_j^{(0)}, \ldots, \boldsymbol{\gamma}_j^{(S)}\}$. To start off, we take the ad hoc estimate of $\boldsymbol{\theta}$ to be the initial value of $\boldsymbol{\theta}^{(0)}$, and we take $\boldsymbol{\gamma}_j^{(0)} = \mathbf{0}$. Therefore, $\mathbf{S}_{\boldsymbol{\gamma}} = \mathbf{0}$ and the starting value of $\Sigma^{(0)}$ is a random sample from Inverse Wishart$(\eta_0 + m, \mathbf{S}_0^{-1})$

Given a state of the parameter vector $\boldsymbol{\gamma}_j^{(s)}$ or $\boldsymbol{\theta}^{(s)}$, a new state of $\boldsymbol{\gamma}_j^{(s+1)}$ or $\boldsymbol{\theta}^{(s+1)}$ is generated using the following steps:

1. Sample a proposal state $\boldsymbol{\gamma}_j^{(*)}$ or $\boldsymbol{\theta}^{(*)}$from multivariate normal $(\boldsymbol{\gamma}_j^{(s)}, k_{\gamma_j}\Sigma^{(s)})$ or $(\boldsymbol{\theta}^{(s)}, k_\theta\Lambda_0)$ depending on which parameter we are proposing. Here, k_{γ_j} and k_θ are scaling parameters to establish suitable variances for sampling the proposed values $\boldsymbol{\gamma}_j^{(*)}$ or $\boldsymbol{\theta}^{(*)}$. Due to the unbalanced data across strata, the variances of the sampling distributions for each stratum are different, requiring different k_{γ_j} value for each stratum j. These scaling parameters, k_{γ_j} and k_θ, are adjusted through trails: If a chosen k results in an excessively low acceptance rate, we decrease k; if it results in an excessively high acceptance rate, we increase k.

2. Calculate the acceptance ratio r

$$r_{\boldsymbol{\gamma}_j} = \frac{\text{dmvnorm}(\boldsymbol{\gamma}_j^{(*)}, \mathbf{0}, \Sigma^{(s)})}{\text{dmvnorm}(\boldsymbol{\gamma}_j^{(s)}, \mathbf{0}, \Sigma^{(s)})} \times \frac{\prod_{i=1}^{n_j} \text{dpois}(Y_{ij} = y_{ij} | \lambda(\mathbf{x}_{ij}, \boldsymbol{\gamma}_j^{(*)}, \boldsymbol{\theta}^{(s)}))^{w_{ij}}}{\prod_{i=1}^{n_j} \text{dpois}(Y_{ij} = y_{ij} | \lambda(\mathbf{x}_{ij}, \boldsymbol{\gamma}_j^{(s)}, \boldsymbol{\theta}^{(s)}))^{w_{ij}}}$$

$$r_{\boldsymbol{\theta}} = \frac{\text{dmvnorm}(\boldsymbol{\theta}^{(*)}, \boldsymbol{\mu}_0, \Lambda_0)}{\text{dmvnorm}(\boldsymbol{\theta}^{(s)}, \boldsymbol{\mu}_0, \Lambda_0)} \times \frac{\prod_{j=1}^{m} \prod_{i=1}^{n_j} \text{dpois}\left(Y_{ij} = y_{ij} | \lambda(\mathbf{x}_{ij}, \boldsymbol{\gamma}_j, \boldsymbol{\theta}^{(*)})\right)^{w_{ij}}}{\prod_{j=1}^{m} \prod_{i=1}^{n_j} \text{dpois}\left(Y_{ij} = y_{ij} | \lambda(\mathbf{x}_{ij}, \boldsymbol{\gamma}_j, \boldsymbol{\theta}^{(s)})\right)^{w_{ij}}}$$

Note that the $\boldsymbol{\gamma}_j$ ratio is calculated over a strata, whereas the $\boldsymbol{\theta}$ ratio is calculated over the whole data sample.

3. Accept or reject the proposal with probability:

$$\boldsymbol{\gamma}_j^{(s+1)} = \begin{cases} \boldsymbol{\gamma}_j^{(*)} & \text{with probability } \min(r, 1) \\ \boldsymbol{\gamma}_j^{(s)} & \text{with probability 1- } \min(r, 1) \end{cases} \tag{15}$$

$$\boldsymbol{\theta}^{(s+1)} = \begin{cases} \boldsymbol{\theta}^{(*)} & \text{with probability } \min(r, 1) \\ \boldsymbol{\theta}^{(s)} & \text{with probability 1- } \min(r, 1) \end{cases} \tag{16}$$

By sampling $u \sim \text{uniform}(0, 1)$.

Set $\boldsymbol{\gamma}_j^{(s+1)} = \boldsymbol{\gamma}_j^{(*)}$ if $u < r$, set $\boldsymbol{\gamma}_j^{(s+1)} = \boldsymbol{\gamma}_j^{(s)}$ if $u \geq r$.

Set $\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(*)}$ if $u < r$, set $\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)}$ if $u \geq r$.

4.5 Process for Sampling The Join Conditional Distributions

Within each sampling iteration, we perform the following steps in order:

1. update Σ by sampling from the posterior of Σ which is the Inverse-Wishart $\left(\eta_0 + m, [\mathbf{S}_0 + \mathbf{S}_{\boldsymbol{\gamma}}]^{-1}\right)$ distribution using Gibbs sampling.

2. Propose a new value for each $\boldsymbol{\gamma}_j$ via Metropolis algorithm, and accept or reject the new value for each strata.

3. Propose a new value for $\boldsymbol{\theta}$ via Metropolis algorithm and accept or reject this new value according to the algorithm.

4.6 Model Evaluation

4.6.1 Model Fitting Evaluation Using BIC

To evaluate the model fitting of the proposed hierarchical mixed effect model, we introduce the concept of Bayesian Information Criterion (BIC) as a metric to measure the model fit on the given dataset.

The BIC, proposed by Gideon Schwarz in 1978, is a widely used tool in statistical modeling for comparing different models and selecting the best one (Schwarz, 1978). It is particularly useful in the context of regression analysis, time-series analysis, and other scenarios where multiple models can be fitted to the same data. By comparing the BIC values of the mixed effect model and the fixed effect model, we can determine which model provides a better fit and more accurately represents the underlying data structure.

For the hierarchical model which incorporate both random and fixed effect, the procedure of calculating the hybrid BIC is defined as follows (Delattre et al., 2014):

$$\begin{aligned}\mathrm{BIC}_h &= -2\log p(\mathbf{y}|\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\gamma}}_j) + \dim(\boldsymbol{\theta}_1)\log m + \dim(\boldsymbol{\theta}_2)\log N \\ &= -2\log p(\mathbf{y}|\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\gamma}}_j) + q\log m + (p-q)\log N\end{aligned}$$

where $\log p(\mathbf{y}|\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\gamma}}_j)$ is the log of the maximum likelihood estimate, and $\hat{\theta}, \hat{\boldsymbol{\gamma}}_j$ are the maximum likelihood estimator (MLE) of fixed and random predictors respectively. `dim(`$\boldsymbol{\theta}$`)` is the number of predictors in $\boldsymbol{\theta}$. m is the number of strata in $\mathbf{y} = (y_1, \ldots, y_N)^T$ and N is the total number of observations.

4.6.2 Model Robustness Evaluation via Changing Priors

In addition to the normal prior specified in previous sections, we verify the robustness of our result by switching to an alternative prior. Three types of prior distributions will be tested: Laplace prior, Cauchy prior and Spike-and-slab prior. (Tomal et al., 2022)

The Laplace prior density function of β_k is defined as follows:

$$\pi(\beta_k|0,\sigma_k) = \frac{1}{2\sigma_k}\exp\left(-\frac{|\beta_k|}{\sigma_k}\right)$$

The Cauchy prior density function of β_k is defined as follows:

$$\pi(\beta_k|0,\sigma_k) = \left[\pi\sigma_k\left(1+\left(\frac{\beta_k}{\sigma_k}\right)^2\right)\right]^{-1}$$

Where the mean (location) parameter of the prior is set to 0 to represent a uninformative prior, and σ_k is the scale parameter, set such that 95% prior probability for each β_k falls within $[-\boldsymbol{\mu}_{0k} - \sigma_k, \boldsymbol{\mu}_{0k} + \sigma_k]$ (recall that $\boldsymbol{\mu}_0$ is an ad-hoc estimation of the regression

coefficients of $\boldsymbol{\theta}$, and $\boldsymbol{\mu}_{0k}$ is the k-th component of the vector $\boldsymbol{\theta}$, $k = 1, 2, \ldots, p$.

For the spike-and-slab prior, we proposed two versions: one discontinuous and the other continuous.

The discontinuous spike-and-slab distribution consists of two components: a Dirac function and a normal function. The corresponding density function is expressed as:

$$f(\theta) = \pi \cdot \delta(\theta) + (1 - \pi) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta - \mu)^2}{2\sigma^2}\right)$$

where:

- π is the mixing proportion, with $\pi = 0$ reducing the distribution to a standard normal prior.
- $\delta(\theta)$ is the Dirac delta function centered at $\theta = 0$.
- μ and σ^2 are the mean and variance of the normal distribution, respectively.

The continuous spike-and-slab prior used in this study is based on Malsiner-Walli and Wagner (2016). This prior combines a "spike" at zero, representing sparsity, with a "slab" that is more dispersed. Both the "spike" and "slab" components are modeled using a Student's t-distribution.

Let $t_\nu(x|\mu, \sigma)$ represent the density function of the Student-t distribution with degrees of freedom ν, location parameter μ, and scale parameter σ.

The density function for the spike-and-slab prior can be written as:

$$\pi(\beta_k) = p_s \cdot t_{\nu_s}\left(\frac{\beta_k}{\sigma_s}\right) \cdot \frac{1}{\sigma_s} + (1 - p_s) \cdot t_{\nu_l}\left(\frac{\beta_k}{\sigma_l}\right) \cdot \frac{1}{\sigma_l}$$

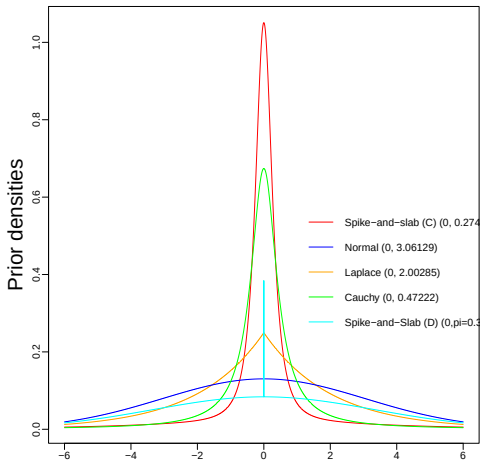


Figure 3: Prior density plot for Normal, Laplace, Cauchy and Spike-and-slab prior. The scale parameter is chosen such that the area under the curve is 95% for a parameter to fall within $[-6, 6]$

The symbols used in the density function are explained in Table 2.

Table 2: Description of the symbols used in the spike-and-slab density function

Symbol	Description
β_j	The variable
ν_s	Degrees of freedom for the spike component
σ_s	Scale (standard deviation) for the spike component
ν_l	Degrees of freedom for the slab component
σ_l	Scale (standard deviation) for the slab component.
p_s	Probability weight for the spike component (where $0 \leq p_s \leq 1$)

An example of the prior set such that 95% prior probability falls within $[-6, 6]$ is plotted in Figure 3.

4.7 Preliminary Result from Bayesian Hierarchical Poisson Model

In this section, we present the preliminary results of our hierarchical mixed effect model and compare them to those obtained from the classic fixed effect Poisson regression model. Table 3 provides the 95% credible intervals (CI) of the regression coefficients for both models. The left three columns display the incidence rate ratios (IRR) and their corresponding credible intervals from the mixed effect model, while the right three columns present the

Table 3: Posterior estimates with 95% credible intervals (CI) of the incidence rate ratio (IRR) for the explannatory variables

Variables	Categories/Type	Mixed effect model			Fixed effect model		
		IRR	95%CI		IRR	95%CI	
			LL	UL		LL	UL
Women's age	Quantitative	1.320	1.319	1.321	1.324	1.318	1.330
Women's age squared	Quantitative	0.997	0.997	0.997	0.997	0.997	0.997
Household head's age	Quantitative	1.005	1.005	1.006	1.004	1.003	1.006
Household head's age squared	Quantitative	1.000	1.000	1.000	1.000	1.000	1.000
Age at marriage	Quantitative	0.958	0.958	0.958	0.957	0.956	0.958
Province	Herat (ref)	...	...	...	...	...	...
	Kabul	0.974	0.969	0.980	0.973	0.951	0.997
	Kapisa	0.967	0.961	0.973	0.962	0.930	1.000
	Parwan	0.990	0.984	0.995	0.995	0.964	1.026
	Maidan Wardak	1.049	1.043	1.055	1.045	1.013	1.077
	Logar	1.070	1.065	1.076	1.066	1.035	1.100
	Nangarhar	1.131	1.125	1.137	1.119	1.091	1.149
	Laghman	1.153	1.146	1.159	1.148	1.117	1.185
	Panjsher	1.016	1.010	1.022	1.013	0.973	1.053
	Baghlan	1.017	1.010	1.023	1.011	0.982	1.042
	Bamyan	0.964	0.959	0.970	0.960	0.929	0.992
	Ghazni	1.008	1.002	1.014	1.008	0.979	1.039
	Paktika	0.936	0.931	0.941	0.930	0.899	0.958
	Paktya	0.985	0.980	0.990	0.983	0.956	1.012
	Khost	1.094	1.088	1.100	1.091	1.065	1.121
	Kunarha	1.182	1.175	1.188	1.174	1.139	1.211
	Nooristan	1.117	1.111	1.123	1.105	1.067	1.141
	Badakhshan	1.075	1.069	1.082	1.053	1.025	1.086
	Takhar	1.081	1.075	1.087	1.082	1.053	1.114
	Kunduz	1.151	1.145	1.157	1.145	1.114	1.178
	Samangan	0.906	0.901	0.912	0.902	0.875	0.935
	Balkh	0.994	0.988	1.000	0.994	0.969	1.021
	Sar-e-Pul	0.949	0.943	0.954	0.943	0.913	0.974
	Ghor	0.991	0.985	0.998	0.984	0.954	1.018
	Daykundi	0.965	0.959	0.971	0.962	0.931	0.991
	Urozgan	1.024	1.018	1.030	1.017	0.983	1.049
	Zabul	1.213	1.206	1.220	1.206	1.171	1.244
	Kandahar	1.232	1.225	1.238	1.221	1.190	1.252
	Jawzjan	0.985	0.980	0.991	0.984	0.955	1.014
	Faryab	1.047	1.041	1.053	1.045	1.017	1.072
	Helmand	1.287	1.280	1.294	1.284	1.256	1.316
	Badghis	1.005	0.999	1.012	0.999	0.968	1.029
	Farah	1.188	1.181	1.195	1.185	1.150	1.222
	Nimroz	1.172	1.165	1.179	1.166	1.127	1.203
Area	Rural (ref)	...	...	...	...	...	...
	Urban	0.997	0.995	1.000	0.995	0.983	1.007
Education of Woman	No(ref)	...	...	...	...	...	...
	Yes	0.938	0.936	0.940	0.929	0.917	0.941
Household head's Edu level	No (ref)	...	...	...	...	...	...
	Yes	0.987	0.985	0.989	0.986	0.977	0.995
Household head's sex	Male (ref)	...	...	...	...	...	...
	Female	0.909	0.905	0.912	0.907	0.889	0.925
Number of other wives	0 (ref)	...	...	...	...	...	...
	≥ 1	0.862	0.859	0.864	0.867	0.856	0.880
Wealth index	Middle (ref)	...	...	...	...	...	...
	Poor	1.029	1.027	1.031	1.026	1.015	1.038
	Rich	0.950	0.948	0.952	0.945	0.933	0.956
Media access	No Access (ref)	...	...	...	...	...	...
	Some Access	0.974	0.972	0.975	0.968	0.959	0.979
BIC		110712.3			111651.9		

results from the fixed effect model (Tomal et al., 2022) for comparison.

The regression analysis results presented in Table 3 offer detailed insights into the factors influencing the incidence rate ratio (IRR) of children ever born (CEB). In this section, we provide a comprehensive interpretation of these results, focusing on each predictor's contribution to the model. By examining the credible intervals and incidence rate ratios, we aim to clarify the relationships between the explanatory variables and fertility rates. The following paragraphs will discuss the effects of women's age, household head's age, women's age at marriage, province of residence, area of residence, education of women, household head's education level, number of other wives, wealth index, and media access on the expected number of CEB. Each explanation is supported by corresponding figures that visualize the predicted outcomes.

The results from the mixed effect model indicate a significant association between women's age and the expected number of children ever born (CEB). As women's age increases, the incidence rate of CEB increases, as shown by the incidence rate ratio (IRR) of 1.320. This indicates that each additional year in women's age is associated with a 32% increase in the incidence rate. The quadratic term (women's age squared) has an IRR of 0.997, suggesting a slight deceleration in the increase of the incidence rate as women's age increases. This non-linear relationship is illustrated in the graph at 7a, which plots the expected CEB against women's age, showing an upward trend that flattens out at older ages.

Household head's age also has a significant effect on the expected CEB. The IRR for household head's age is 1.005, indicating a 0.5% increase in the incidence rate for each additional year. The quadratic term for household head's age has an IRR of 1.000, indicating no significant non-linear effect. This relationship is depicted in the graph at 8a, which plots the expected CEB against household head's age, showing a slight upward trend without substantial curvature.

Women's age at marriage is another significant predictor of CEB. The IRR for age at marriage is 0.958, indicating that each additional year in age at marriage is associated with a 4.2% decrease in the incidence rate. This suggests that women who marry later tend to have fewer children. The graph at 9a illustrates this relationship, plotting the expected CEB against women's age at marriage and showing a downward trend.

The results for the fixed predictor Province show significant variation in the expected CEB across different provinces. For example, compared to Herat (the reference province), provinces like Kabul and Kapisa have lower incidence rates, while provinces like Nangarhar and Kunarha have higher incidence rates. This spatial variation in CEB is visualized in the map at Figure 12a, which shows different expected CEB values for each province, highlighting regional differences in fertility patterns.

The area of residence (Urban vs. Rural) is also a significant predictor. Living in urban areas is associated with a slightly lower incidence rate (IRR: 0.997), indicating a 0.3% decrease compared to rural areas. The expected CEB for urban and rural areas is plotted in 10a, showing a slightly lower CEB in urban areas compared to rural areas.

Education of women is another important factor. Women with education have a lower incidence rate (IRR: 0.938), indicating a 6.2% decrease compared to those without education. This suggests that educated women tend to have fewer children. The graph at 13a plots the expected CEB against women's education status, showing a lower expected CEB for educated women.

Household head's education level also influences CEB. Households where the head has an education have a lower incidence rate (IRR: 0.987), indicating a 1.3% decrease compared to those where the head does not have an education. The expected CEB for households with and without educated heads is plotted in 14a, showing a lower expected CEB for households with educated heads.

The number of other wives in the household is another significant predictor. Households with one or more additional wives have a lower incidence rate (IRR: 0.862), indicating a 13.8% decrease compared to those with no additional wives. This suggests that in polygamous households, the number of children per wife is lower. The graph at 16a plots the expected CEB against the number of other wives, showing a lower expected CEB for households with additional wives.

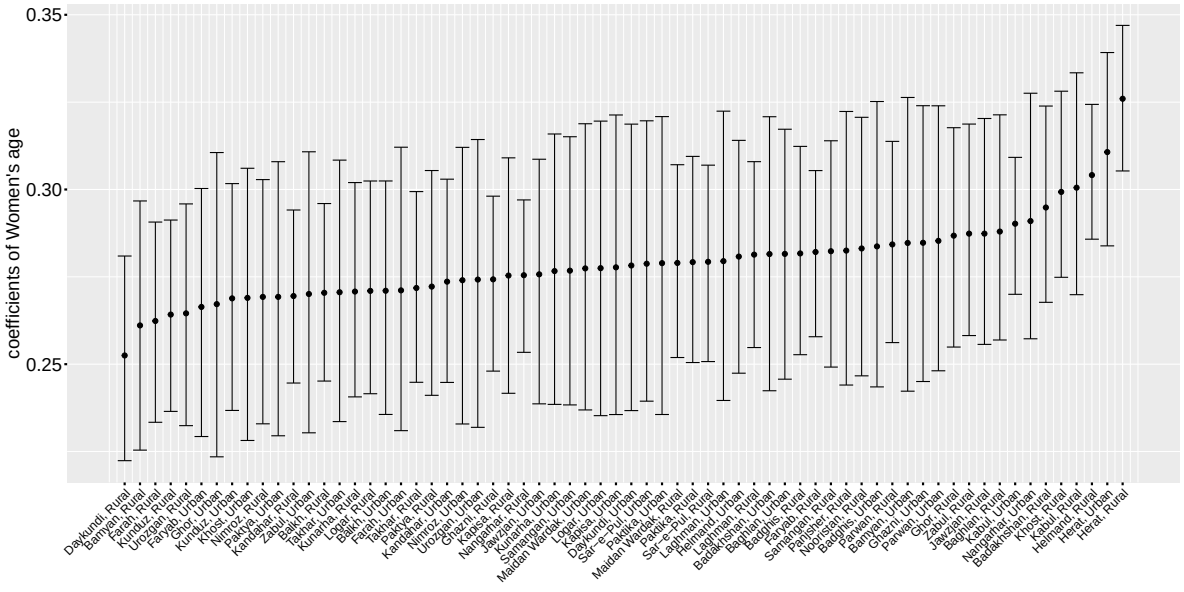
The wealth index is also a significant predictor. Poor households have a higher incidence rate (IRR: 1.029), indicating a 2.9% increase compared to middle wealth households, while rich households have a lower incidence rate (IRR: 0.950), indicating a 5.0% decrease. This suggests that wealthier households tend to have fewer children. The expected CEB for different wealth levels is plotted in 17a, showing a higher expected CEB for poor households and a lower expected CEB for rich households.

Media access is another important factor. Households with some media access have a lower incidence rate (IRR: 0.974), indicating a 2.6% decrease compared to those with no media access. This suggests that access to media is associated with lower fertility rates. The graph at 18a plots the expected CEB against media access status, showing a lower expected CEB for households with some media access.

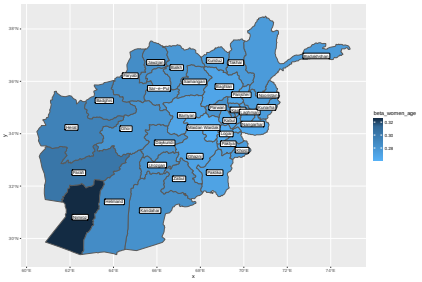
In conclusion, the mixed effect model results highlight the significant influence of various demographic, socio-economic, and regional factors on fertility rates in Afghanistan. These findings can inform policy interventions aimed at addressing population dynamics and promoting reproductive health.

4.7.1 Evaluating random effects across strata

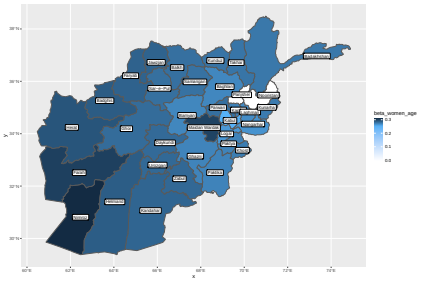
We applied a hierarchical Bayesian model to examine the effects of random predictors on the number of children ever born (CEB) across 66 strata, representing both urban and rural areas within each province. This approach enables us to discern differences in how these



(a) 95% credible intervals for the regression coefficients of women's age in each stratum



(b) This map shows the regression coefficients of women's age in rural areas



(c) This map shows the regression coefficients of women's age in urban areas

Figure 4: 95% Credible interval for the regression coefficients of women's age in each stratum and geographic plots

predictors influence fertility across different regions.

In our model, we included random coefficients $\boldsymbol{\gamma}_j$ for certain predictors, which vary by strata. These random coefficients are combined with the fixed effect $\boldsymbol{\theta}$ to obtain $\boldsymbol{\beta}_j$, the regression coefficient specific to each predictor and stratum. Predictors marked with an "R" in Table 1 are those for which random effects were considered.

The predictor "women's age" is found to be significant across all strata. The 95% credible intervals for the regression coefficients of women's age in each stratum, along with geographic plots, are displayed in Figure 4.

The positive coefficient of women's age indicates that, as a woman's age increases, the expected number of children ever born (CEB) also increases. This suggests a general trend where older women are likely to have more children.

The negative quadratic coefficient implies that the rate of increase in the expected CEB diminishes with age. After reaching a certain age, the expected CEB might start to decline across the entire population of women.

When comparing the geographic distribution of the random coefficients for "women's age" between the rural and urban area of each province, we observe notable regional differences. Specifically, provinces in the western part of the country show a stronger tendency for women to have more children as they age, compared to other regions.

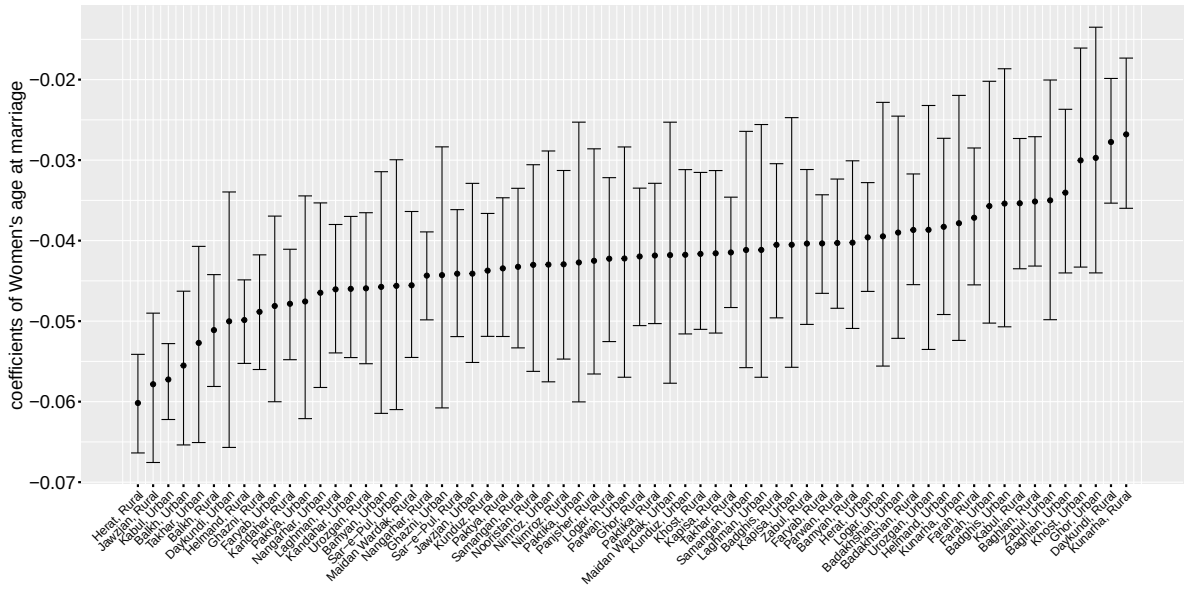
The predictor "women's age at marriage" also shows significance across all strata. The 95% credible intervals for the regression coefficients of women's age at marriage in each stratum, along with geographic plots, are displayed in Figure 5.

The negative coefficient of women's age at marriage indicates that, as a woman's age at marriage increases, the expected number of children ever born (CEB) decreases, indicating that earlier marriage is associated with higher fertility.

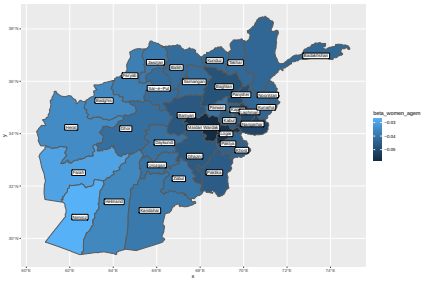
When comparing the geographic distribution of the random coefficients for "women's age at marriage" between the rural and urban area of each province, we observe notable regional differences. Specifically, Kabul and its neighbouring provinces show a stronger tendency for women to have less children if the are married at a older age, compared to other regions.

For the rest random predictors, although not all of them are significant across all strata, regional differences are still evident.

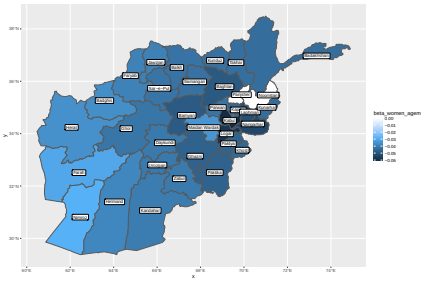
For women's education, in the northeastern part of the country, having an education significantly reduces the number of children ever born. This effect is especially evident in the



(a) 95% credible intervals for the regression coefficients of women's age at marriage in each stratum

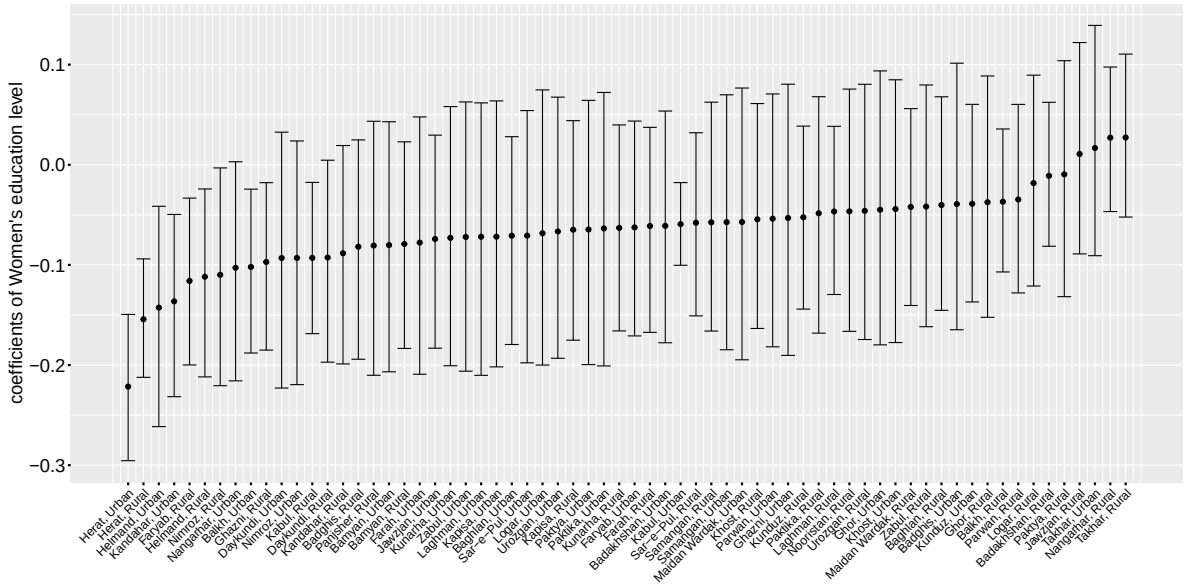


(b) This map shows the regression coefficients of women's age at marriage in rural areas

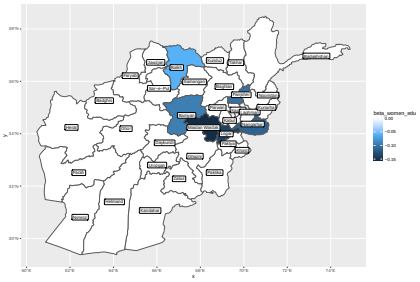


(c) This map shows the regression coefficients of women's age at marriage in urban areas

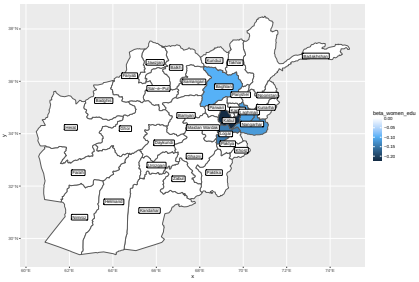
Figure 5: 95% Credible interval for the regression coefficients of women's age at marriage in each stratum and geographic plots



(a) 95% credible intervals for the regression coefficients of women's education in each stratum



(b) This map shows the regression coefficients of women's education in rural areas

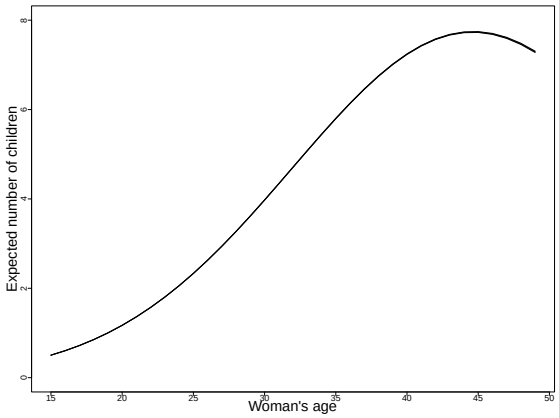


(c) This map shows the regression coefficients of women's education in urban areas

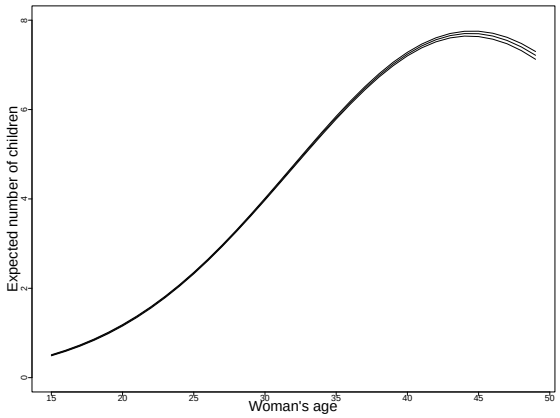
Figure 6: 95% Credible interval for the regression coefficients of women's education in each stratum and geographic plots

rural areas of Maidan Wardak, Logar, Nangarhar, Bamyan, Panjsher, and Balkh, as well as in the urban areas of Kabul, Logar, Nangarhar, Laghman, and Baghlan. The 95% credible intervals for the regression coefficients of women's age at marriage in each stratum, along with geographic plots, are displayed in Figure 6.

These patterns suggests a significant geographic variation in fertility behavior, potentially influenced by cultural, economic, or social factors specific to these areas.

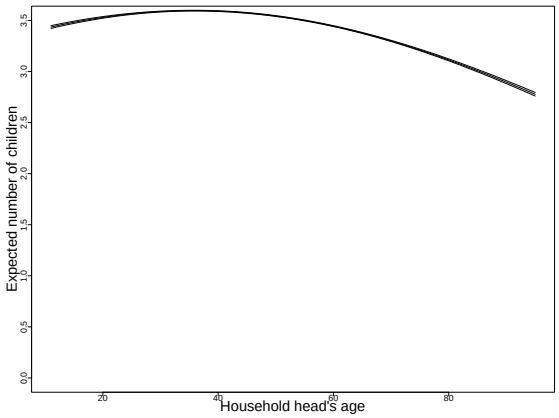


(a) Mixed effect model: Expected CEB - Women's age

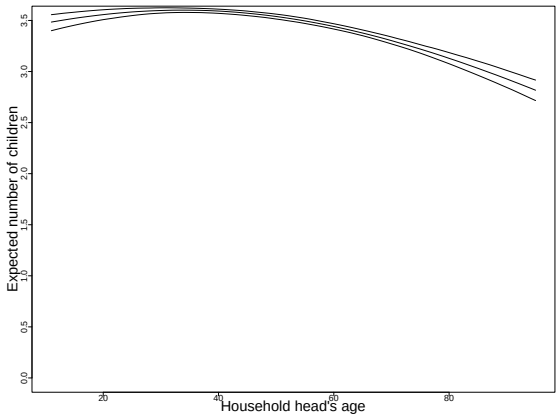


(b) Fixed effect model: Expected CEB - Women's age

Figure 7: Estimates of expected CEB against women's age, Fixed effect model and mixed effect model

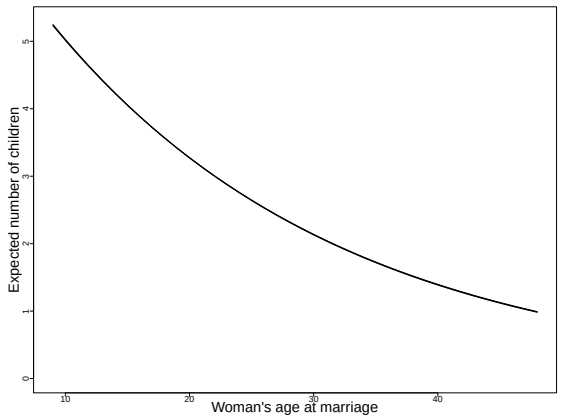


(a) Mixed effect model: Expected CEB - Household head's age

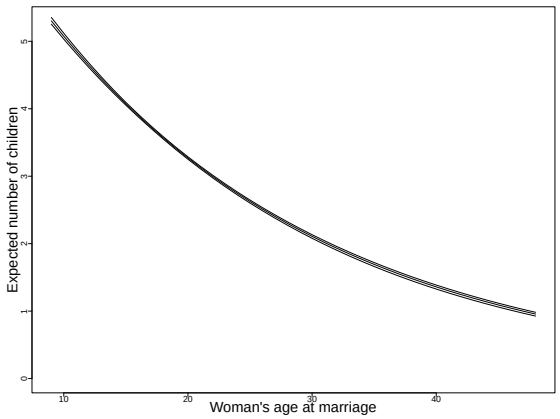


(b) Fixed effect model: Expected CEB - Household head's age

Figure 8: Estimates of expected CEB against Household head's age, Fixed effect model and mixed effect model

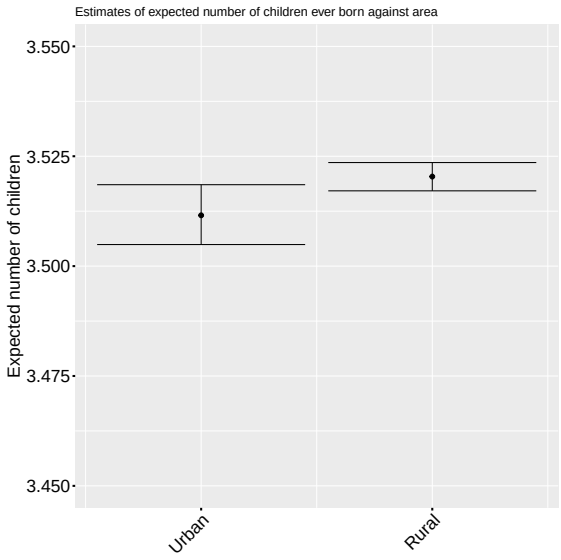


(a) Mixed effect model: Expected CEB - Women's age at marriage

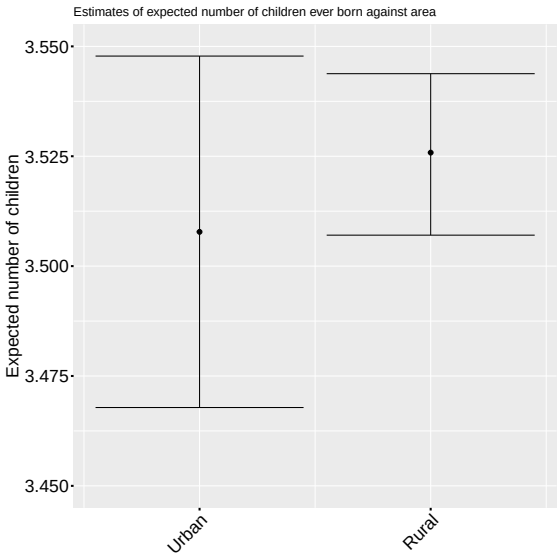


(b) Fixed effect model: Expected CEB - Women's age at marriage

Figure 9: Estimates of expected CEB against women's age at marriage, Fixed effect model and mixed effect model

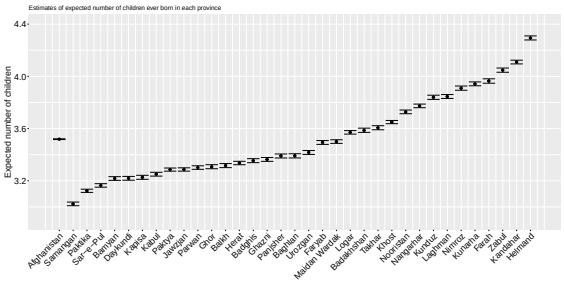


(a) Mixed effect model: Expected CEB - Area

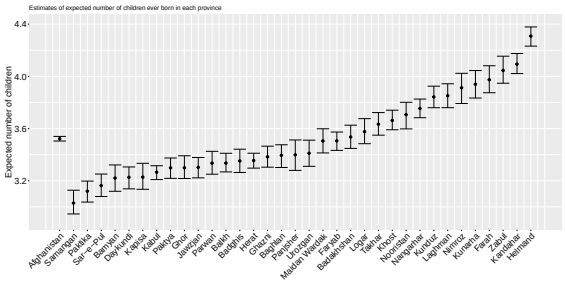


(b) Fixed effect model: Expected CEB - Area

Figure 10: Estimates of expected CEB against area, Fixed effect model and mixed effect model

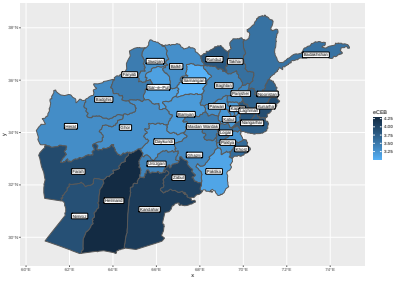


(a) Mixed effect model: Expected CEB - Province

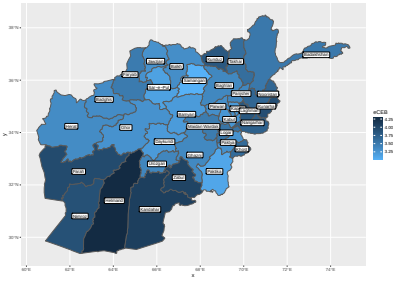


(b) Fixed effect model: Expected CEB - Province

Figure 11: Estimates of expected CEB against province, Fixed effect model and mixed effect model

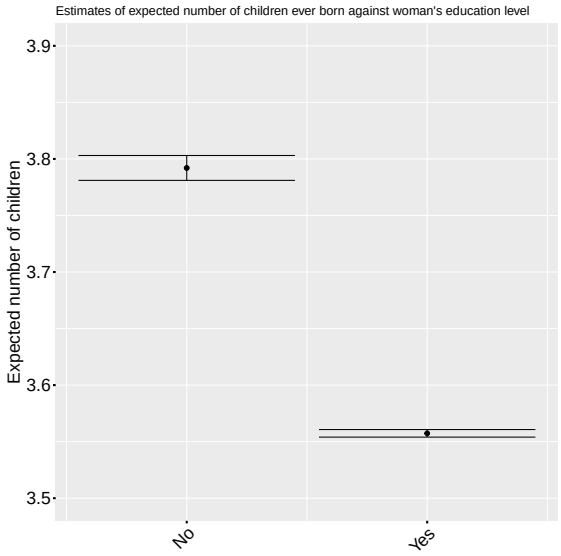


(a) Mixed effect model: Expected CEB - Province

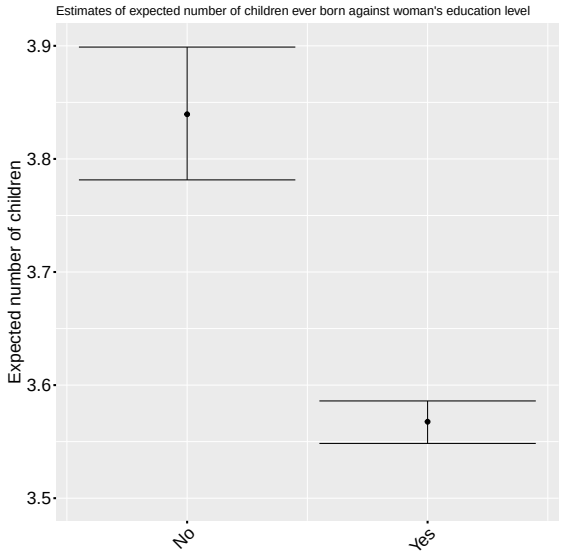


(b) Fixed effect model: Expected CEB - Province

Figure 12: Geographic plots of Estimated Expected CEB by Province: Comparison of Fixed Effect and Mixed Effect Models

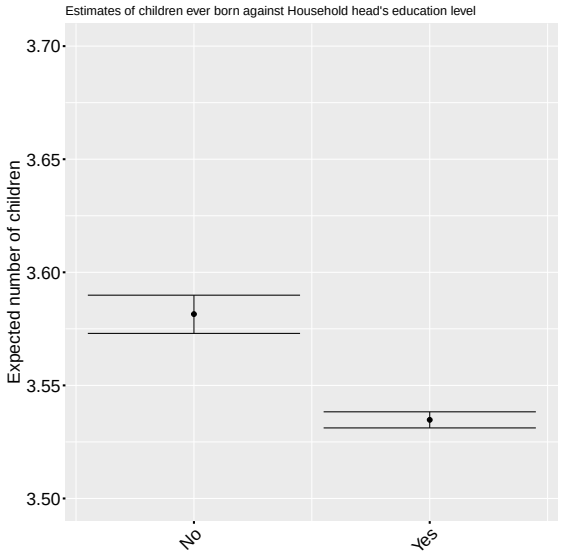


(a) Mixed effect model: Expected CEB - Women's education

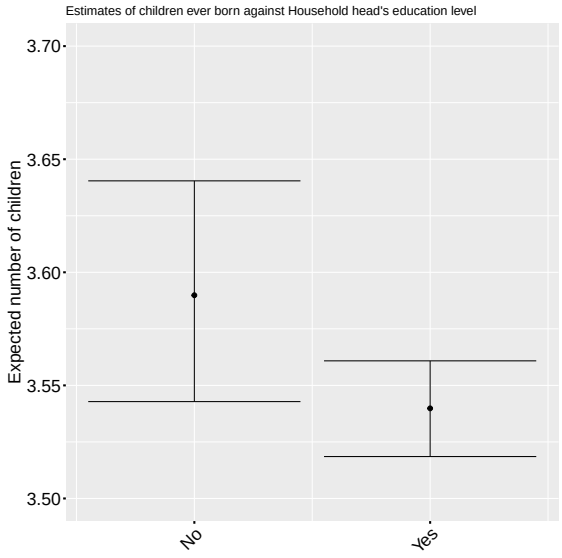


(b) Fixed effect model: Expected CEB - Women's education

Figure 13: Estimates of expected CEB against women's education, Fixed effect model and mixed effect model

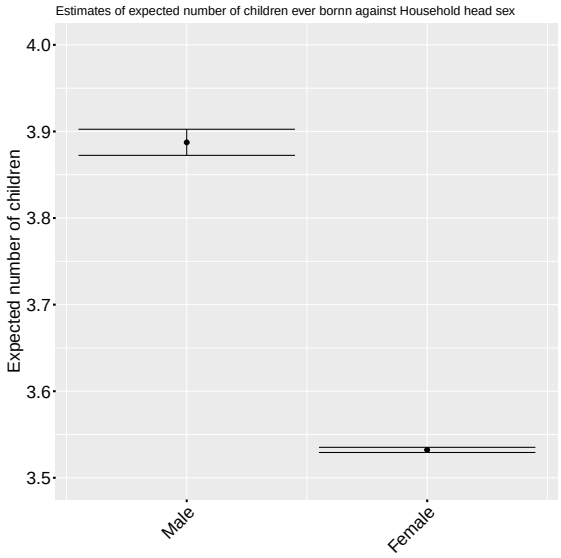


(a) Mixed effect model: Expected CEB - Household head's education

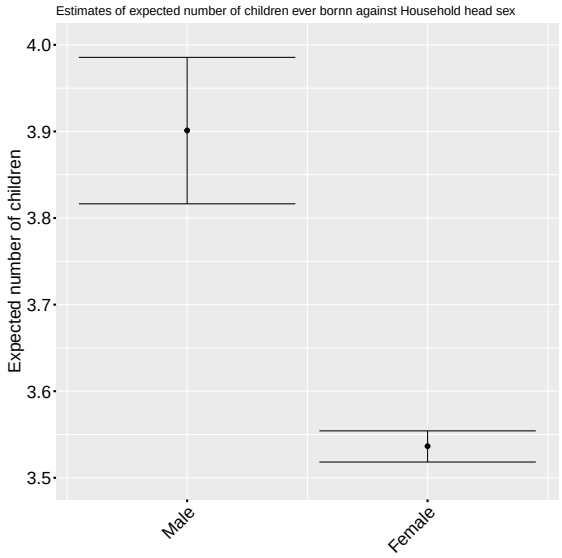


(b) Fixed effect model: Expected CEB - Household head's education

Figure 14: Estimates of expected CEB against household head's education, Fixed effect model and mixed effect model

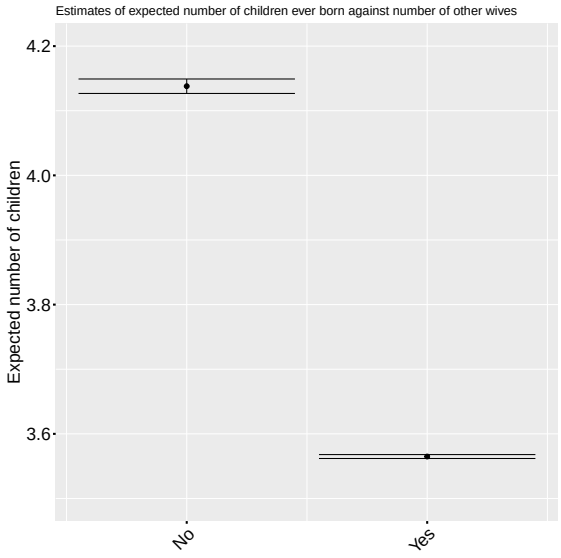


(a) Mixed effect model: Expected CEB - Household head's sex

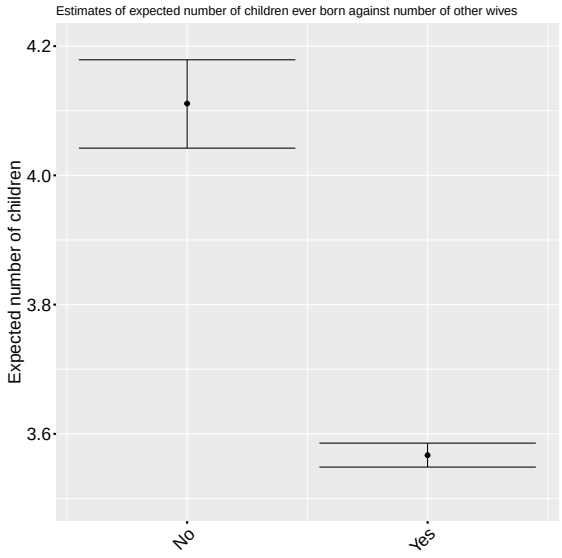


(b) Fixed effect model: Expected CEB - Household head's sex

Figure 15: Estimates of expected CEB against household head's sex, Fixed effect model and mixed effect model

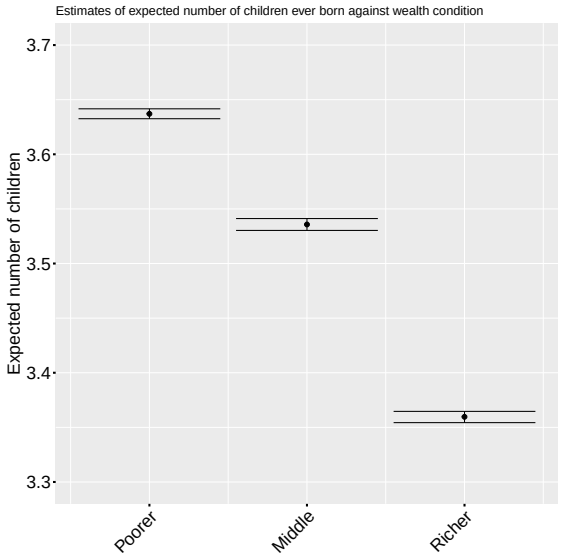


(a) Mixed effect model: Expected CEB - Having other wives

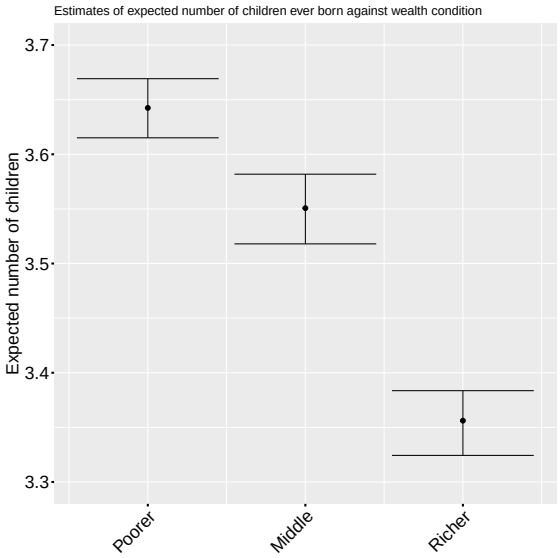


(b) Fixed effect model: Expected CEB - Having other wives

Figure 16: Estimates of expected CEB against Number of other wives, Fixed effect model and mixed effect model

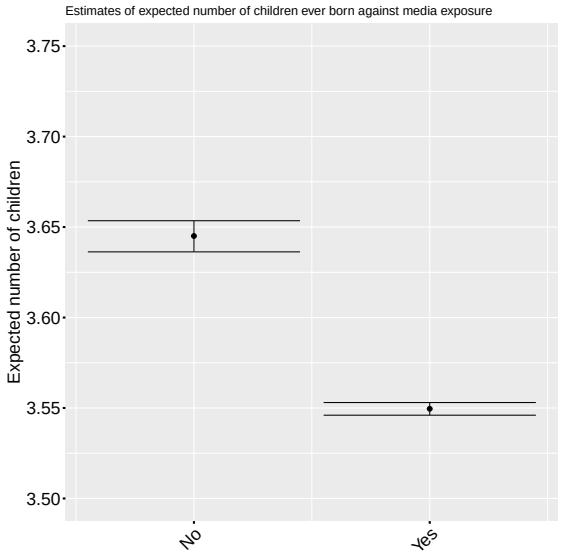


(a) Mixed effect model: Expected CEB - Wealth index

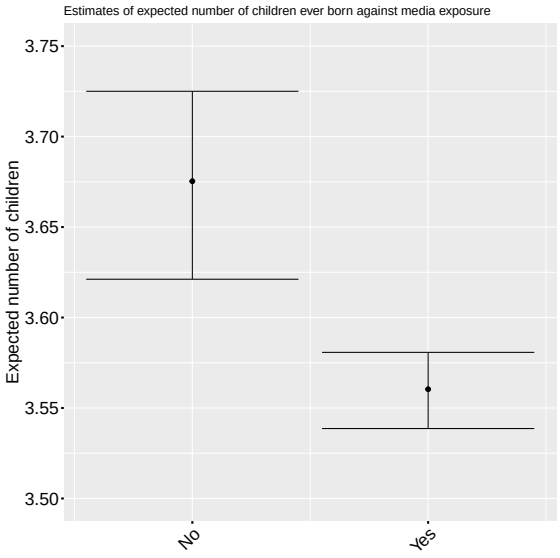


(b) Fixed effect model: Expected CEB - Wealth index

Figure 17: Estimates of expected CEB against Wealth index, Fixed effect model and mixed effect model



(a) Mixed effect model: Expected CEB - media access



(b) Fixed effect model: Expected CEB - media access

Figure 18: Estimates of expected CEB against Wealth index, Fixed effect model and mixed effect model

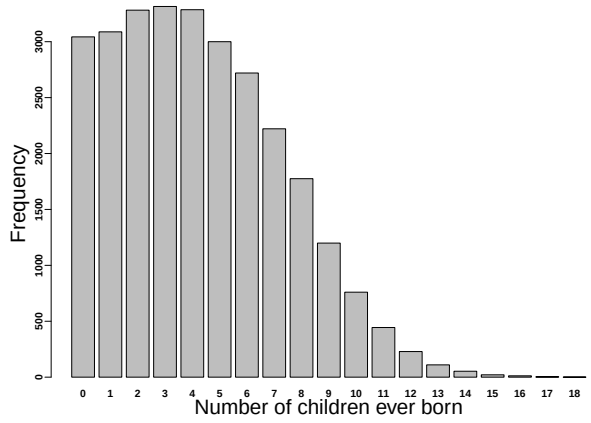


Figure 19: Bar-plot showing the distribution of the number of children ever born to married women of reproductive age in Afghanistan

4.8 Dispersion Test

Figure 19 shows the distribution of CEB which closely resembles a Poisson probability distribution. The `dispersiontest()` function from the `AER` package in R is used to assess the equidispersion assumption in Poisson Generalized Linear Models (GLMs). This assumption states that the mean and variance of the dependent variable y are equal, i.e.,

$$E[y] = \mu \quad \text{and} \quad \text{Var}[y] = \mu.$$

The function tests for the presence of overdispersion or underdispersion in the data by comparing the observed dispersion with the expected dispersion under the Poisson distribution.

The null hypothesis of equidispersion assumes that the variance of y is equal to the mean:

$$H_0 : \text{Var}[y] = \mu.$$

The alternative hypothesis tests for either overdispersion or underdispersion, where the variance takes the form:

$$\text{Var}[y] = \mu + \alpha \cdot \text{trafo}(\mu),$$

where α is a coefficient and `trafo` is a transformation of the mean μ. Common choices for `trafo` are:

- trafo(μ) = μ: Linear variance function (quasi-Poisson or NB1 (Cameron and Trivedi, 2005) model), used by default.
- trafo(μ) = μ^2: Quadratic variance function (NB2 (Cameron and Trivedi, 2005) model).

Overdispersion occurs when $\alpha > 0$, meaning the variance is greater than the mean, while underdispersion corresponds to $\alpha < 0$, meaning the variance is less than the mean. If the null hypothesis is rejected, it implies that the equidispersion assumption does not hold.

The test statistic for the `dispersiontest()` is based on an auxiliary Ordinary Least Squares (OLS) regression, and the resulting z-statistic follows a standard normal distribution under the null hypothesis of equidispersion.

We use a Classic Poisson GLM model to fit the data and ran the dispersion test with the alternative hypothesis of underdispersion (`alternative = "less"`). The test yielded the following results:

```
Underdispersion test
data:  fit.mle
z = -17.054, p-value < 2.2e-16
alternative hypothesis: true dispersion is less than 1
sample estimates:
dispersion
 0.8569994
```

The z-value of -17.054 is highly significant, as indicated by the p-value $< 2.2 \times 10^{-16}$. This suggests strong evidence against the null hypothesis of equidispersion. Since the test was performed with the alternative hypothesis of underdispersion, the results indicate that the

data exhibits underdispersion, where the variance is smaller than the mean. The estimated dispersion parameter is approximately 0.857, meaning that the variance is about 85.7% of the mean.

The results of the dispersion test indicate that the classic Poisson model, which assumes equidispersion, is inadequate for this dataset. With a significantly negative z-value and an estimated dispersion parameter less than one, we observe underdispersion in the data, where the variance is smaller than the mean. This finding suggests that using the Poisson distribution might not be sufficient to accurately model the fertility data. Since the Poisson model cannot accommodate the underdispersion identified in our analysis, a more flexible distribution is necessary. The Conway-Maxwell-Poisson (COM-Poisson) regression model offers such flexibility by introducing an additional parameter, ν, which adjusts the variance independently of the mean. This model is particularly well-suited for count data exhibiting underdispersion or overdispersion, and will help address the limitations found in the classic Poisson model.

5 Proposed Extension: COM-Poisson Model

Extending the model from a classic Poisson to a COM-Poisson introduces an additional parameter, ν. The goal is to derive the posterior distribution for this new parameter and to sample it from the corresponding distribution.

5.0.1 The Sampling Distribution

The likelihood for a set of n identically distributed observations $y_1, y_2, \ldots, y_n$ is

$$
\begin{aligned}
L(y_1, y_2, \ldots, y_n|\lambda, \nu) &= \frac{\prod_{i=1}^{n} \lambda^{y_i}}{(\prod_{i=1}^{n} y_i!)^{\nu}} Z^{-n}(\lambda, \nu) \\
&= \lambda^{\sum_{i=1}^{n} x_i} e^{-\nu \sum_{i=1}^{n} \log(x_i!)} Z^{-n}(\lambda, \nu) \\
&= \lambda^{S_1} e^{-\nu S_2} Z^{-n}(\lambda, \nu)
\end{aligned} \tag{17}
$$

where $S_1 = \sum_{i=1}^{n} x_i$ and $S_2 = \sum_{i=1}^{n} \log(x_i!)$. By the Factorization Theorem, (S_1, S_2) are sufficient statistics. Furthermore, Eq.17 displays the COM-Poisson as an exponential family of distributions.

5.0.2 Prior Distribution

Since the COM-Poisson forms an exponential family, there is a conjugate family of priors such that, whatever the data, the posterior is of the same form. For this distribution, the conjugate prior density is of the form

$$
h(\lambda, \nu) = \lambda^{a-1} e^{-\nu b} Z^{-c}(\lambda, \nu) \kappa(a, b, c) \tag{18}
$$

for $\lambda > 0$ and $\nu \geq 0$, where $\kappa(a, b, c)$ is the integration constant. Then the posterior is of the same form, with $a^{'} = a + S_1, b^{'} = b + S_2$, and $c^{'} = c + n$. The distribution whose density is given in Eq. 18 can be thought of as an extended bivariate Gamma distribution.

In order for Eq. 18 to constitute a density, it must be non-negative and integrate to one. In other words, the values of a, b and c that lead to a finite $\kappa^{-1}(a, b, c)$, which is given by

$$
\kappa^{-1}(a, b, c) = \int_0^{\infty} \int_0^{\infty} \lambda^{a-1} e^{-b\nu} Z^{-c}(\lambda, \nu) d\lambda d\nu \tag{19}
$$

will lead to a proper density.

Using Jensen's inequality and the convexity of the log-gamma function, we can show that a necessary and sufficient condition for a finite $\kappa^{-1}(a, b, c)$ is:

$$b/c > \log(\lfloor a/c \rfloor!) + (a/c - \lfloor a/c \rfloor) \log(\lfloor a/c \rfloor + 1) \tag{20}$$

where $\lfloor k \rfloor$ denotes the floor of k. The set of hyperparameters (a, b, c) satisfying Eq. 20 is necessarily closed under sampling; i.e., if the prior satisfies Eq. 20, so will the posterior for every possible datum x.

6 Timeline (draft)

May 01: Started thesis

Aug 01: First Committee meeting

Aug 29: Proposal Defence

Sep 05: [Thesis only] Meeting with supervisor to create timeline for defence, (e.g., minimum 3 months before anticipated defence date, when first draft to supervisor, final draft to external/university no later than 1.5 months before defence, etc.).

Sep 05: Fit a GLM model using the unmerged categorical predictors to evaluate whether a more fine-grained approach reveals additional insights..

Sep 12: Implement the extension of COM-Poisson Model.

Sep 19: Compare the results with the current model and provide dive explanation and insights into the new findings.

Sep 30: first complete draft

Oct 30: second draft

Nov 20: last day to submit final draft to circulate to committee/second reader

Dec 20: presentation/defence (determined by coordinator)

Thesis students should also budget thesis supervisory meetings once a term at a time con-

venient for all members. E.g., end of the first month of a term (Sept 30 and Jan 30).

7 Supervisory Dissolution

The student agrees that supervision will be dissolved if any of the following happen:

- Two consecutive progress reports are unacceptable
- Three consecutive progress reports are concerning/unacceptable
- An academic integrity violation is suspected by the supervisor and suspected by at least one other faculty member.

8 Outcomes and attribution

The expected outcomes of this study will provide both practical and theoretical contributions to the field of fertility analysis, particularly within the context of Afghanistan. By developing a Bayesian hierarchical weighted COM-Poisson regression model, this research will offer a more flexible approach to understanding fertility patterns, effectively addressing the issues of overdispersion and underdispersion that have been challenging to handle using traditional Poisson models.

Here are the expected details of outcomes:

- Enhanced Model Accuracy: The proposed model is expected to offer a more precise representation of fertility dynamics, improving upon existing models by accounting for the variance discrepancies inherent in real-world fertility data. This will help produce more accurate estimates of the number of children ever born (CEB) and the factors influencing these outcomes across different regions of Afghanistan.
- Identification of Key Predictors: The study will highlight the significant individual,

household, and regional factors that influence fertility in Afghanistan. Specifically, the research will shed light on how variables such as education, age, wealth, media exposure, and geographic location impact fertility, helping policymakers tailor interventions more effectively.

- Regional Insights: The study will offer an in-depth understanding of regional disparities in fertility patterns. By capturing differences between urban and rural areas and across provinces, this research will support efforts to address regional inequalities in family planning, reproductive health services, and resource allocation.

- Model Robustness: By evaluating model fit through the Bayesian Information Criterion (BIC) and testing robustness under different prior distributions, this research will demonstrate the superiority of the COM-Poisson model in handling complex fertility data. The comparative analysis with previous models will further validate the strength of the proposed approach.

Attribution:

The findings from this research will be directly attributable to the innovative application of the hierarchical COM-Poisson regression model, a novel approach in the field of demographic analysis. This work will build on the foundation of existing literature but will push the boundaries by providing a more nuanced tool for fertility analysis in Afghanistan, particularly in the context of conflict-affected regions where fertility trends are often shaped by complex socio-economic and cultural factors.

Additionally, the success of this study can be attributed to the comprehensive use of recent data from the UNICEF MICS dataset, which allows for a more detailed and current examination of fertility patterns. The inclusion of regional differences and the handling of both overdispersion and underdispersion make this research an important contribution to demographic and health-related policymaking.

Ultimately, the outcomes of this research will serve as a resource for future studies on fertility in Afghanistan and similar contexts, and will assist governments, NGOs, and healthcare providers in developing more effective programs that address the unique needs of various population groups, leading to improved health and social outcomes.

References

Abbasi-Shavazi, M., G. (dec, R. Sadeghi, and H. Mahmoudian (2015, 07). Immigrant-native fertility differentials: The afghans in iran. *Asian and Pacific Migration Journal 24*.

Aslam, K., S. Zaheer, M. S. Qureshi, S. N. Aslam, and K. Shafique (2016). Socio-economic disparities in use of family planning methods among pakistani women: Findings from pakistan demographic and health surveys. *PloS one 11*(4), e0153313.

Britten, G. L., Y. Mohajerani, L. Primeau, M. Aydin, C. Garcia, W.-L. Wang, B. Pasquier, B. B. Cael, and F. W. Primeau (2021). Evaluating the benefits of bayesian hierarchical methods for analyzing heterogeneous environmental datasets: A case study of marine organic carbon fluxes. *Frontiers in Environmental Science 9*.

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in r. *Advances in Methods and Practices in Psychological Science 4*(1), 2515245920960351.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Central Statistics Organization (CSO), M. o. P. H. M. (2017). Afghanistan demographic and health survey 2015. Dataset and report. Available at: `https://microdata.worldbank.org/index.php/catalog/2786`.

Conway, R. W. and W. L. Maxwell (1962). A queueing model with state dependent service rate. *Journal of Industrial Engineering 12*, 132–136.

Dadras, O., T. Khampaya, and T. Nakayama (2022). Child marriage, reproductive outcomes, and service utilization among young afghan women: Findings from a nationally representative survey in afghanistan. *Studies in Family Planning 53*(3), 417–431.

Delattre, M., M. Lavielle, and M.-A. Poursat (2014). A note on BIC in mixed-effects models. *Electronic Journal of Statistics 8*(1), 456 – 475.

Ehsan, H., N. Ghafoori, N. Ghafoori, H. Ehsan, S. O. Akrami, and S. O. Akrami (2021). The impact of poverty and education on female child marriage in afghanistan evidence from 2015 afghanistan demographic and health survey. *19 Mayıs Sosyal Bilimler Dergisi*.

Erkan Ozkaya, H., C. Dabas, K. Kolev, G. T. M. Hult, S. H. Dahlquist, and S. A. Manjeshwar (2013). An assessment of hierarchical linear modeling in international business, management, and marketing. *International Business Review 22*(4), 663–677.

Fagbamigbe, A. F. and A. S. Adebowale (2014). Current and predicted fertility using poisson regression model: Evidence from 2008 nigerian demographic health survey. *African Journal of Reproductive Health / La Revue Africaine de la Santé Reproductive 18*(1), 71–83.

Hosseini-Chavoshi, M., M. Abbasi-Shavazi, and P. Mcdonald (2017, 01). Fertility, marriage, and family planning in iran: Implications for future policy. *Population Horizons 13*.

Indian Institute for Health Management Research (IIHMR), C. S. O. C. (2013). Afghanistan mortality survey 2010. Dataset and report. Available at: `https://microdata.worldbank.org/index.php/catalog/1322`.

Malsiner-Walli, G. and H. Wagner (2016). Comparing spike and slab priors for bayesian

variable selection. *Austrian Journal of Statistics 40*, 241–264.

Osmani, A., J. Reyer, A. Osmani, and N. Hamajima (2015, Nov). Factors influencing contraceptive use among women in afghanistan: secondary analysis of afghanistan health survey 2012. *Nagoya Journal of Medical Science 77*(4), 551–561.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Shams Ghahfarokhi, M. (2024). Investigating the relationship between spousal violence against women and total fertility rate in afghanistan. *BMC Public Health 24*(1463).

Tomal, J. H., J. H. Tomal, J. R. Khan, J. R. Khan, A. S. Wahed, and A. S. Wahed (2022). Weighted bayesian poisson regression for the number of children ever born per woman in bangladesh. *Journal of Statistical Theory and Applications*.

UNICEF (2023). Afghanistan multiple indicator cluster survey 2022-23: Summary findings report.

9 Appendix

9.1 Proof of Λ_0

Proof:

Since $(Y_{11}, \ldots, Y_{n_m m}) \overset{\text{renumber}}{=} (Y_1, \ldots, Y_N)$[1] are independent Poisson random variables, the likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{t=1}^{N} \frac{\lambda_t^{Y_t} e^{-\lambda_t}}{Y_t!} \tag{21}$$

where $\lambda_t = \exp(\mathbf{x}_t \boldsymbol{\theta})$ The log-likelihood is:

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{N} Y_t \log(\lambda_t) - \lambda_t - \log(Y_t!) \tag{22}$$

$$= \sum_{t=1}^{N} Y_t \left(\mathbf{x}_t \boldsymbol{\theta}\right) - \exp(\mathbf{x}_t \boldsymbol{\theta}) - \log(Y_t!) \tag{23}$$

and the MLEs are the solutions to the system of score equations, for $u = 1, \ldots, p$

$$0 = \frac{\partial l}{\partial \boldsymbol{\theta}_u} = \sum_{t=1}^{N} x_{t,u}(Y_t - \exp(\mathbf{x}_t \boldsymbol{\theta})) \tag{24}$$

where $x_{t,u}$ is the u-th element in vector $\mathbf{x}_t$.

The Fisher information matrix $I_{\mathbf{Y}}(\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{\theta}}[\nabla^2 l(\boldsymbol{\theta})]$ may be obtained by computing the second-order partial derivatives of l:

[1] In our hierarchical model, we have data grouped into m strata, where each stratum j contains n_j observations. Originally, we denote each observation with a double subscript Y_{ij}, where i indicates the observation within the j-th stratum. To simplify notation, we renumber these observations using a single subscript from 1 to N, where N is the total number of observations across all strata.

$$\frac{\partial^2 l}{\partial \boldsymbol{\theta}_u \partial \boldsymbol{\theta}_v} = -\sum_{t=1}^{N} x_{t,u} x_{t,v} \exp(\mathbf{x}_t \boldsymbol{\theta}) \tag{25}$$

Using the definition in Eq. 10, the above may be written as $\frac{\partial^2 l}{\partial \boldsymbol{\theta}_u \partial \boldsymbol{\theta}_v} = -\mathbf{X}_{[u]}^{\mathrm{T}} \mathbf{W} \mathbf{X}_{[v]}$, [2] so $\nabla^2 l(\boldsymbol{\theta}) = -\mathbf{X}^{\mathrm{T}} \mathbf{W} \mathbf{X}$ and $I_{\mathbf{Y}}(\boldsymbol{\theta}) = \mathbf{X}^{\mathrm{T}} \mathbf{W} \mathbf{X}$. For large N, if Poisson log-linear model is correct, then the MLE vector $\hat{\boldsymbol{\theta}}$ is approximately distributed as $\mathrm{MVN}(\boldsymbol{\theta}, (\mathbf{X}^{\mathrm{T}} \mathbf{W} \mathbf{X})^{-1})$ ■

9.2 Derive for posterior distribution

9.2.1 Sampling distribution:

$$\begin{aligned}
&P(\mathbf{Y} = \mathbf{y}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m | \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}, \Sigma) \\
=&P(\mathbf{Y} = \mathbf{y} | \mathbf{Z}, \mathbf{X}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}) \times P(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m | \mathbf{0}, \Sigma) \\
=&\prod_{j=1}^{m} \left[\mathrm{dmvnorm}(\boldsymbol{\gamma}_j, \mathbf{0}, \Sigma) \times \prod_{i=1}^{n_j} \mathrm{dpois}(Y_{ij} = y_{ij} | \exp\left(\mathbf{z}_{ij} \boldsymbol{\gamma}_j + \mathbf{x}_{ij} \boldsymbol{\theta}\right))^{w_{ij}} \right] \\
=&\prod_{j=1}^{m} \left[(2\pi)^{-q/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2} \boldsymbol{\gamma}_j^{\mathrm{T}} \Sigma^{-1} \boldsymbol{\gamma}_j\} \times \prod_{i=1}^{n_j} \left(\frac{e^{-\exp(\mathbf{z}_{ij} \boldsymbol{\gamma}_j + \mathbf{x}_{ij} \boldsymbol{\theta})} [\exp(\mathbf{z}_{ij} \boldsymbol{\gamma}_j + \mathbf{x}_{ij} \boldsymbol{\theta})]^{y_{ij}}}{y_{ij}!} \right)^{w_{ij}} \right]
\end{aligned}$$

[2] Here $\mathbf{X}_{[u]}, \mathbf{X}_{[v]}$ is the u, v column of the design matrix $\mathbf{X}$

9.2.2 Joint conditional distribution:

$$
\begin{aligned}
&P(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}, \Sigma | \mathbf{y}, \mathbf{Z}, \mathbf{X}) \\
\propto & P(\mathbf{Y} = \mathbf{y}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m | \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}, \Sigma) \times P(\boldsymbol{\theta}) \times P(\Sigma) \\
= & \prod_{j=1}^{m} \left[(2\pi)^{-q/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2} \boldsymbol{\gamma}_j^{\mathrm{T}} \Sigma^{-1} \boldsymbol{\gamma}_j\} \right. \\
& \left. \times \prod_{i=1}^{n_j} \left(\frac{e^{-\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})} [\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})]^{y_{ij}}}{y_{ij}!} \right)^{w_{ij}} \right] \times P(\boldsymbol{\theta}) \times P(\Sigma) \\
= & \prod_{j=1}^{m} \left[(2\pi)^{-q/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2} \boldsymbol{\gamma}_j^{\mathrm{T}} \Sigma^{-1} \boldsymbol{\gamma}_j\} \right. \\
& \left. \times \prod_{i=1}^{n_j} \left(\frac{e^{-\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})} [\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})]^{y_{ij}}}{y_{ij}!} \right)^{w_{ij}} \right] \\
& \times (2\pi)^{-p/2} |\Lambda_0|^{-1/2} \exp\{-\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu}_0)^{\mathrm{T}} \Lambda_0 (\boldsymbol{\theta} - \boldsymbol{\mu}_0)\} \\
& \times \left[2^{\eta_0 p/2} \pi^{\binom{p}{2}/2} |\mathbf{S}_0|^{-\eta_0/2} \prod_{t=1}^{p} \Gamma([\eta_0 + 1 - t]/2) \right]^{-1} \times |\Sigma|^{-(\eta_0 + p + 1)/2} \times \exp\{-\mathrm{tr}(\mathbf{S}_0 \Sigma^{-1})/2\}
\end{aligned}
$$

9.2.3 Full conditional distribution of $\boldsymbol{\theta}$

$$
\begin{aligned}
& P(\boldsymbol{\theta} | \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m, \Sigma, \mathbf{y}, \mathbf{Z}, \mathbf{X}) \\
\propto & \prod_{j=1}^{m} \prod_{i=1}^{n_j} \left(\frac{e^{-\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})} [\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})]^{y_{ij}}}{y_{ij}!} \right)^{w_{ij}} \\
& \times (2\pi)^{-p/2} |\Lambda_0|^{-1/2} \exp\{-\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu}_0)^{\mathrm{T}} \Lambda_0 (\boldsymbol{\theta} - \boldsymbol{\mu}_0)\}
\end{aligned}
$$

9.2.4 Full conditional distribution of $\boldsymbol{\gamma}_j$

$$P(\boldsymbol{\gamma}_j|\boldsymbol{\theta}, \Sigma, \mathbf{y}_j, \mathbf{Z}_j, \mathbf{X}_j) \tag{26}$$

$$\propto P(\mathbf{Y}_j = \mathbf{y}_j|\mathbf{Z}_j, \mathbf{X}_j, \boldsymbol{\gamma}_j, \boldsymbol{\theta}) \times P(\boldsymbol{\gamma}_j|\mathbf{0}, \Sigma) \tag{27}$$

$$= \prod_{i=1}^{n_j} \text{dpois}(Y_{ij} = y_{ij}| \exp\left(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta}\right))^{w_{ij}} \times \text{dmvnorm}(\boldsymbol{\gamma}_j, \mathbf{0}, \Sigma) \tag{28}$$

$$= \prod_{i=1}^{n_j} \left(\frac{e^{-\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})}[\exp(\mathbf{z}_{ij}\boldsymbol{\gamma}_j + \mathbf{x}_{ij}\boldsymbol{\theta})]^{y_{ij}}}{y_{ij}!} \right)^{w_{ij}} \times (2\pi)^{-q/2}|\Sigma|^{-1/2} \exp\{-\frac{1}{2}\boldsymbol{\gamma}_j^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\gamma}_j\} \tag{29}$$

$$\tag{30}$$

9.2.5 Full conditional distribution of Σ

$$P(\Sigma|\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m, \boldsymbol{\theta}, \mathbf{y}, \mathbf{Z}, \mathbf{X}) \tag{31}$$

$$\propto P(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m|\mathbf{0}, \Sigma) \times P(\Sigma) \tag{32}$$

$$\propto \prod_{j=1}^{m} |\Sigma|^{-1/2} \exp(-\frac{1}{2}\boldsymbol{\gamma}_j^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\gamma}_j) \times |\Sigma|^{-(\eta_0+q+1)/2} \exp\left(-\text{tr}(\mathbf{S}_0\Sigma^{-1})/2\right) \tag{33}$$

$$\propto |\Sigma|^{-(\eta_0+m+q+1)/2} \exp\left(-\text{tr}((\mathbf{S}_0 + \mathbf{S}_{\boldsymbol{\gamma}})\Sigma^{-1})/2\right) \tag{34}$$

where $S_{\boldsymbol{\gamma}} = \sum_{j=1}^{m}(\boldsymbol{\gamma}_j)(\boldsymbol{\gamma}_j)^{\mathrm{T}}$ is the sum of squares of the deviations of $\boldsymbol{\gamma}_j$ from $\mathbf{0}$.

9.3 Conway-Maxwell-Poisson distribution

Proof that Eq. 20 is a Necessary Condition ($\Rightarrow$):

A lower bound on $\kappa^{-1}(a, b, c)$ is obtained via an upper bound on $Z(\lambda, \nu)$, which in turn comes from a lower bound on the factorial function. Because the log-gamma function is convex, i.e. has a positive second derivative, we can lower bound the log-factorial with a linear function. In particular, the following bound is valid for all integers j and x:

$$\log j! \geq \log x! + (j-x)\log(x+1) \tag{35}$$

The bound is tight at $j = x$ and $j = x+1$. This gives the following bound $\overline{Z}(\lambda,\nu)$ on $Z(\lambda,\nu)$:

$$Z(\lambda,\nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!^\nu} \leq \sum_{j=0}^{\infty} \frac{\lambda^j}{x!^\nu (x+1)^{(j-x)\nu}} = \overline{Z}(\lambda,\nu) \tag{36}$$

$$\overline{Z}(\lambda,\nu) = \frac{(x+1)^{x\nu}}{x!^\nu} \sum_{j=0}^{\infty} \left(\frac{\lambda}{(x+1)^\nu} \right)^j \tag{37}$$

$$= \begin{cases} \frac{(x+1)^{x\nu}}{x!^\nu} \frac{1}{1-\lambda(x+1)^{-\nu}}, & \text{if } \lambda < (x+1)^\nu \\ \infty & \text{otherwise} \end{cases} \tag{38}$$

From this we obtain a lower bound on the double integral:

$$\kappa^{-1}(a,b,c) \geq \int_0^\infty e^{-b\nu} \int_0^\infty \frac{\lambda^{a-1}}{\overline{Z}(\lambda,\nu)^c} d\lambda d\nu \tag{39}$$

$$= \int_0^\infty \frac{e^{-b\nu} x!^{c\nu} (x+1)^{a\nu}}{(x+1)^{cx\nu}} \int_0^{(x+1)^\nu} \lambda^{a-1} (1-\lambda(x+1)^{-\nu})^c d\lambda d\nu \tag{40}$$

Change from λ to $\omega = \lambda(x+1)^{-\nu}$:

$$\kappa^{-1(a,b,c)} \geq \int_0^\infty \frac{e^{-b\nu} x!^{c\nu} (x+1)^{a\nu}}{(x+1)^{cx\nu}} d\nu \int_0^1 \omega^{a-1} (1-\omega)^c d\omega \tag{41}$$

The integral over ω is always finite. The integral over ν is finite only if

$$b/c > \log x! + (a/c - x)\log(x+1) \tag{42}$$

Proof that Eq. 20 is a Sufficient Condition ($\Leftarrow$):

An upper bound on $\kappa^{-1}(a,b,c)$ is obtained by breaking the integral into two parts and bounding each part:

$$\kappa^{-1}(a,b,c) = \int_0^\infty e^{-b\nu} \int_0^1 \frac{\lambda^{a-1}}{Z(\lambda,\nu)^c} d\lambda d\nu + \int_0^\infty e^{-b\nu} \int_1^\infty \frac{\lambda^{a-1}}{Z(\lambda,\nu)^c} d\lambda d\nu \tag{43}$$

$$= I_1 + I_2 \tag{44}$$

Since $Z(\lambda,\nu) \geq 1$, $I_1 \leq \int_0^\infty e^{-b\nu} \int_0^1 \lambda^{a-1} d\lambda d\nu$, which is finite for all $a > 0$ and $b > 0$. We use a different lower bound on $Z(\lambda,\nu)$ to handle I_2. Because $\log(x)$ is concave, i.e. the second derivative is always negative, we know from Jensen's inequality that

$$\log(\sum_{j=0}^\infty q_j a_j) \geq \sum_{j=0}^\infty q_j \log a_j \quad \text{if} \sum_{j=0}^\infty q_j = 1 \tag{45}$$

Therefore by introducing variables q_j we have

$$\log Z(\lambda,\nu) = \log \sum_{j=0}^\infty q_j \frac{\lambda^j}{q_j (j!)^\nu} \geq \sum_{j=0}^\infty q_j \log\left(\frac{\lambda^j}{q_j (j!)^\nu}\right) \tag{46}$$

$$= \left(\sum_{j=0}^\infty j q_j\right) \log\lambda - \nu \left(\sum_{j=0}^\infty q_j \log(j!)\right) - \sum_{j=0}^\infty q_j \log q_j \tag{47}$$

Let Q be a random variable on the non-negative integers with probability mass function

$Pr(Q = j) = q_j$. Then the bound can be written succinctly as

$$\underline{Z}(\lambda, \nu) = \lambda^{E[Q]} e^{-E[\log Q!]\nu} \prod_{j=0}^{\infty} q_j^{q_j} \tag{48}$$

Now we have an upper bound on the double integral:

$$I_2 \leq \int_0^{\infty} e^{-b\nu} \int_1^{\infty} \frac{\lambda^{a-1}}{\underline{Z}(\lambda, \nu)^c} d\lambda d\nu \tag{49}$$

$$= \int_0^{\infty} e^{-b\nu} e^{cE[\log(Q!)]\nu} d\nu \int_1^{\infty} \lambda^{a-1} \lambda^{-cE(Q)} d\lambda \prod_{j=0}^{\infty} q_j^{-cq_j} \tag{50}$$

This integral and therefore $\kappa^{-1}(a, b, c)$ is finite if

$$E(Q) > a/c \text{ and } E(\log Q!) < b/c \tag{51}$$

Given Eq. 20, we just have to show that there exists a distribution satisfying Eq. 51. Let

$$q_j = \begin{cases} 1 - (a/c - \lfloor a/c \rfloor + \epsilon), & \text{if } j = \lfloor a/c \rfloor \\ a/c - \lfloor a/c \rfloor + \epsilon, & \text{if } j = \lfloor a/c \rfloor + 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then $E[Q] = a/c + \epsilon$ and $E[\log Q!] = \log \lfloor a/c! \rfloor + (a/c - \lfloor a/c \rfloor + \epsilon) \log(\lfloor a/c \rfloor + 1)$