Journal of Statistical Models and Methods (2026)
2(1), 36–48, https://doi.org/10.65905/zbapb565

AdvantaCore Scientific Publishing

Article Type: Research article

A Probabilistic Model for Household Labor Income and Consumption: Theory and Application with Saudi HICES Microdata

Ateq Ahmed AlGhamedi

Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding Author
Ateq Ahmed AlGhamedi
ateq@kau.edu.sa

Article History

Received: 26 January 2026
Revised: 04 March 2026
Accepted: 27 April 2026

Keywords
Hierarchical probabilistic modeling; Compound count–lognormal distributions; Poisson thinning; Stochastic Engel curves; Survey-weighted microsimulation; Counterfactual policy evaluation

Abstract

This study proposes a hierarchical probabilistic model for household labor income and consumption, designed for microsimulation and policy analysis. Labor income is modeled as a compound count–lognormal sum, arising from zero-truncated household size, Bernoulli labor participation, and i.i.d. lognormal wages, yielding a point mass at zero and a heavy right tail. Consumption follows a log-linear Engel curve in total resources, capturing sub-unitary elasticity and heteroskedasticity. The model provides closed-form cumulants and moments, supports lognormal-sum approximations, and integrates survey-weighted estimation aligned with Saudi HICES reference periods and design. Simulations confirm accurate parameter estimation, while a Universal Basic Income (UBI) counterfactual shows nonlinear inequality reductions, stronger for consumption due to Engel curve concavity, with decomposition highlighting the distinct roles of participation and wage dispersion. The framework is extensible and policy-ready.

1 Introduction

Micro-founded modeling of household income and consumption is central to welfare measurement, tax–benefit analysis, and inequality research. At the micro level, outcomes arise from the interplay of discrete mechanisms, such as household composition and labor force participation, and continuous mechanisms, including individual wage dispersion and the nonlinear translation of resources into consumption. Together, these generate empirical regularities that resist representation by purely continuous models: structural zeros (e.g., jobless households), pronounced right-skewness in income, and concave, heteroskedastic Engel curves.

This paper proposes an analytically tractable generative model that explicitly decomposes four core components: (i) household size, (ii) labor participation, (iii) wage dispersion, and (iv) consumption behavior. Household size follows a zero-truncated count distribution; labor participation is modeled via independent Bernoulli trials, yielding a thinned count of earners; individual earnings are i.i.d. lognormal, producing household labor income as a compound count–lognormal sum with a point mass at zero and a heavy right tail. Consumption is linked to total resources, labor income plus non-labor transfers, through a log-linear Engel curve, generating sub-unitary elasticity and heteroskedastic lognormal errors consistent with empirical evidence.

The model is designed for microsimulation: it admits closed-form cumulants and raw moments for income, supports approximation methods for sums of lognormals (e.g., the Fenton–Wilkinson technique), and facilitates Monte Carlo implementation, Fenton [1]. A principal contribution is its explicit integration

Copyright © 2026 by the author. Published under Creative Commons Attribution (CC BY) license.

with Saudi Arabia's Household Income and Consumption Expenditure Survey (HICES). The HICES methodology specifies that income and inter-household transfers refer to the 12 months preceding the interview, while diary-based daily expenditures cover the 14 days following the first visit, and other spending categories have reference periods of one month, three months, or 12 months, depending on the category, General Authority for Statistics (GASTAT) [2]. This study aligns the model's timing structure with these heterogeneous measurement windows and implements design-consistent, survey-weighted estimation and inference.

Recent advances in flexible probabilistic and parametric distributions provide an important methodological foundation for modeling socioeconomic microdata characterized by heavy tails, structural zeros, truncation, and complex dependence. Hassan et al. [3] extended the exponentiated Pareto family, a class well suited to highly skewed outcomes, and developed estimation procedures that improve inference under substantial dispersion. Similarly, Kamal et al. [4] proposed an improved Lomax-based model with a bivariate extension, showing how flexible distributional forms can capture dependence across related economic variables while also supporting inequality-oriented analysis through tools such as Lorenz and Bonferroni curves. The importance of truncation in probabilistic modeling is further emphasized by Ahmadini et al. [5], who developed the truncated Lomax–inverse Lomax distribution and demonstrated how truncated and mixture-based constructions can better accommodate complex data structures than standard models. In a related direction, Mahmoud et al. [6] underscored the value of generalized Beta-type extensions for representing heterogeneity and inequality in empirical distributions. More broadly, Tang et al. [7] and Kamal et al. [4] also illustrated the growing emphasis on analytically tractable probabilistic models supported by Monte Carlo evidence, reinforcing the importance of combining theoretical derivation with simulation-based validation. Building on these contributions, the present study extends the literature by developing a hierarchical probabilistic framework for household labor income and consumption in which labor income arises from a compound count–lognormal process driven by zero-truncated household size, Bernoulli labor participation, and lognormal wages, thereby generating both a point mass at zero and a heavy right tail. Consumption is modeled through a log-linear Engel curve in total resources, allowing for sub-unitary elasticity and heteroskedasticity. Unlike earlier studies, which focus mainly on abstract distributional innovation or applications in finance, reliability, biomedical, and related fields, the current framework is explicitly designed for survey-weighted microsimulation and policy analysis using Saudi HICES data. Its contribution lies not only in its flexible stochastic structure, but also in its derivation of closed-form moments and cumulants, its use of tractable lognormal-sum approximations, and its direct applicability to counterfactual policy experiments such as universal basic income simulations, where nonlinear reductions in inequality can be quantified separately through participation and wage-dispersion channels.

The specific research gap addressed in this paper is, therefore, the absence of a unified, analytically tractable, and survey-compatible framework that jointly models household composition, labor-force participation, wage dispersion, and nonlinear consumption behavior for microsimulation and policy evaluation in the Saudi context.

Beyond filling this gap, the model is motivated by a practical empirical need: policy analysts require a framework that can distinguish whether inequality is primarily driven by demographic exposure to work, participation failures, wage heterogeneity among workers, or the nonlinear translation of resources into consumption. Existing studies rarely bring these channels together in a single structure that remains transparent enough for decomposition and tractable enough for simulation-based counterfactuals.

An empirical illustration using simulated data demonstrates accurate parameter recovery, robust inequality diagnostics, and decomposition of inequality into participation and wage components. Extensions accommodate intra-household dependence and dynamic behavioral responses, while robustness diagnostics validate key assumptions. The framework synthesizes insights from statistical size distributions, Kleiber and Kotz [8], inequality measurement theory, Cowell [9] and Atkinson [10], consumption survey methodology, Deaton [11], and compound stochastic processes, Willmot and Lin [12], offering a statistically rigorous and policy-relevant platform for distributional analysis and counterfactual evaluation of fiscal interventions such as Universal Basic Income.

The remainder of the paper is organized as follows. Section 2 summarizes the measurement structure of Saudi HICES and its implications for modeling. Sections 3–8 develop the probabilistic framework and its key distributional properties. Section 9 discusses estimation and calibration, Section 10 presents the policy counterfactual, and the final sections discuss limitations, conclusions, and future extensions.

2 Saudi HICES: Measurement, Reference Periods, and Microdata Use

Saudi HICES (Household Income and Expenditure Survey) is a nationally representative household survey covering the Kingdom's 13 administrative regions, GASTAT [2]. The methodology report details reference periods: income and transfers are measured over the 12 months preceding the visit; daily food and related spending is recorded via diaries over the 14 days following the first visit; and other expenditure categories use specific recall periods (1 month, 3 months, or 12 months), GASTAT [2]. Data collection involves multiple visits within a 14-day window, GASTAT [2].

The 2023 publication reports a target sample of 122,325 households and provides monthly averages for disposable income and consumption expenditure, GASTAT [13]. HICES outputs are disseminated as monthly indicators, so empirical implementation should harmonize all modeled flows to a monthly unit (e.g., annual income divided by 12; 14-day diaries scaled to monthly totals using the survey's standardization rules).

Microdata access. The methodology report describes microdata access pathways (public use vs. scientific use) and emphasizes anonymization, GASTAT [2]. Researchers typically obtain Scientific Use Files via GASTAT's approved channels and must comply with confidentiality requirements.

Implication for modeling. Because disposable income in HICES includes multiple sources, labor income S and non-labor resources R are distinguished, with total resources defined as $Y = S + R$. Consumption C is modeled as a function of Y (or $S + \delta$ when R is unobserved or absorbed into δ).

3 Model Specification

This section specifies a generative microsimulation framework for household labor income and consumption that explicitly accounts for demographic structure, labor market participation, wage heterogeneity, and nonlinear consumption behavior. All random variables are assumed to be mutually independent unless otherwise stated, an assumption that ensures analytical tractability while preserving the core dimensions of heterogeneity observed in empirical data. Dependence across components can be flexibly introduced during estimation through covariates or hierarchical structures, without compromising the model's foundational clarity.

Each component serves a distinct economic purpose. Household size determines the potential number of earners, participation governs whether that potential is realized, wage dispersion shapes the upper tail of labor income, and the Engel curve translates resources into welfare-relevant consumption outcomes. Presenting the model in this layered way is useful not only mathematically but also substantively, as it makes subsequent policy decomposition more interpretable.

3.1 Household Size and Labor Participation

Let N denote the number of individuals in a non-empty household. To exclude degenerate (empty) households while retaining flexibility, N follows a zero-truncated Poisson distribution:

$$\mathbb{P}(N = n \mid N \geq 1) = \frac{e^{-\lambda}\lambda^n/n!}{1 - e^{-\lambda}}, \quad n = 1, 2, \ldots, \quad \lambda > 0. \tag{1}$$

The conditional mean is

$$\mathbb{E}[N \mid N \geq 1] = \frac{\lambda}{1 - e^{-\lambda}}, \tag{2}$$

which exceeds the untruncated Poisson mean λ. If empirical dispersion exceeds the Poisson benchmark, a zero-truncated negative binomial distribution can be substituted with minimal downstream changes.

Conditional on N, each member participates independently in the labor market:

$$W_j \sim \text{Bernoulli}(p), \quad j = 1, \ldots, N, \quad 0 < p < 1, \tag{3}$$

where $W_j = 1$ indicates positive labor income. The total number of earners is

$$M \mid N \sim \text{Binomial}(N, p). \tag{4}$$

Thus, $M = 0$ occurs with positive probability, naturally generating a point mass at zero in labor income, a feature essential for realism.

3.2 Participation, Number of Earners, and Household Labor Income

The formulation above yields jobless households ($M = 0$) with positive probability. To accommodate intra-household dependence or unobserved heterogeneity in participation propensity, the model can be extended to a beta–binomial specification, where p varies across households according to a Beta(α, β) distribution.

Given $M \geq 1$, individual labor earnings are modeled as i.i.d. lognormal random variables:

$$Z_k \sim \text{Lognormal}(\mu, \sigma^2), \quad \text{i.e.,} \quad \log Z_k \sim \mathcal{N}(\mu, \sigma^2), \quad k = 1, \ldots, M.$$

The lognormal assumption captures pronounced right-skewness and heavy tails in empirical wage distributions while retaining analytical tractability.

Total household labor income S is defined as

$$S = \sum_{k=1}^{M} Z_k, \quad \text{with} \quad S = 0, \quad \text{if} \quad M = 0. \tag{5}$$

Thus, S follows a compound count–lognormal distribution, a mixed discrete–continuous law featuring: (i) a point mass at zero (from non-participation), (ii) a heavy right tail (from lognormal wages), and (iii) overdispersion driven by randomness in both the number and earnings of workers.

This structure aligns closely with empirical regularities in income data, Kleiber and Kotz [8]. If wage correlation within households is warranted, the i.i.d. assumption can be relaxed via a multivariate lognormal distribution.

3.3 Non-labor Resources and Consumption

HICES disposable income includes non-labor components (transfers, remittances, etc.). Total monthly resources are

$$Y = S + R, \quad R \geq 0, \tag{6}$$

where R can be directly measured from HICES or modeled as a separate positive distribution (e.g., lognormal or gamma). If R is not used explicitly, a positive shift δ serves as a reduced-form proxy.

Consumption follows a log-linear Engel curve:

$$\log C = \beta_0 + \beta_1 \log(Y + \delta) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \tau^2), \tag{7}$$

with $\beta_1 \in (0, 1)$ to encode concavity (declining marginal propensity to consume). Heteroskedasticity can be incorporated by allowing τ^2 to depend on Y.

4 Distribution of the Number of Earners M

The random variable M, representing the number of income earners in a household, plays a pivotal role in our model: it determines whether total labor income is zero or positive and governs the degree of aggregation in the compound income process. Its distribution arises from a two-stage stochastic mechanism, first, household size is drawn; second, each member independently decides to work. This structure leads to a well-known result in probability theory known as Poisson thinning, which we now formalize and extend to the empirically relevant case of non-empty households.

4.1 Untruncated Benchmark and Poisson Thinning

Proposition 1 (Poisson thinning). If $N \sim \text{Poisson}(\lambda)$ (untruncated) and $M \mid N \sim \text{Binomial}(N, p)$, then

$$M \sim \text{Poisson}(\nu), \quad \nu = \lambda p.$$

This benchmark result is useful because it yields a simple closed-form representation for the expected number of earners, $\nu = \lambda p$. Economically, it shows that the labor-income process is driven jointly by demographic scale, how many potential earners live in the household, and by labor-market attachment, the probability that each potential earner actually works.

4.2 Observed Households and Truncation $N \geq 1$

Surveys observe non-empty households. Under latent $N \sim \text{Poisson}(\lambda)$, but conditioning on $N \geq 1$, the distribution of $M \mid N \geq 1$ is zero-modified Poisson:

$$\mathbb{P}(M = 0 \mid N \geq 1) = \frac{e^{-\nu} - e^{-\lambda}}{1 - e^{-\lambda}}, \tag{8}$$

and for $m \geq 1$,

$$\mathbb{P}(M = m \mid N \geq 1) = \frac{e^{-\nu}\nu^m/m!}{1 - e^{-\lambda}}. \tag{9}$$

Ignoring truncation understates the jobless-household share when λ is small and $p < 1$.

This correction matters for empirical welfare analysis because the lower tail of the household income distribution is particularly sensitive to the frequency of zero-earner households. Even modest underestimation of this mass can distort inequality measures and attenuate the simulated effect of cash-transfer policies targeted at low-resource households.

5 Distribution of Labor Income S

Total household labor income S is a central object in our model: it drives consumption, determines tax liability, and underpins measures of inequality and poverty. Because S aggregates stochastic earnings across a random number of workers, its distribution inherits complexity from both the count process (number of earners) and the continuous wage distribution. This leads to a rich mixture structure that captures key empirical features, most notably, a point mass at zero and a highly skewed positive component.

5.1 Mixture Structure

Conditional on $M = m$, income is a sum of m lognormals:

$$S \mid M = m = \sum_{k=1}^{m} Z_k, \quad m = 0, 1, 2, \ldots$$

Thus S has a point mass at zero and a continuous component over $(0, \infty)$:

$$\mathbb{P}(S = 0) = \mathbb{P}(M = 0), \quad f_S(s) = \sum_{m \geq 1} \mathbb{P}(M = m) f_{S_m}(s), \quad s > 0,$$

where $S_m = \sum_{k=1}^{m} Z_k$. For $m \geq 2$, f_{S_m} has no closed form.

5.2 Exact Cumulants and Moments

Let $m_r = \mathbb{E}[Z^r]$. For a compound Poisson sum under the untruncated benchmark $M \sim \text{Poisson}(\nu)$,

$$\kappa_r(S) = \nu m_r, \quad r = 1, 2, 3, \ldots \tag{10}$$

In particular,

$$\mathbb{E}[S] = \kappa_1 = \nu m_1, \quad \text{Var}(S) = \kappa_2 = \nu m_2, \quad \mu_3(S) = \kappa_3 = \nu m_3.$$

Raw moments are obtained from cumulants; for example,

$$\mathbb{E}[S^3] = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3 = \nu m_3 + 3(\nu m_2)(\nu m_1) + (\nu m_1)^3. \tag{11}$$

For lognormal $Z \sim \text{Lognormal}(\mu, \sigma^2)$,

$$m_r = \exp\left(r\mu + \frac{r^2\sigma^2}{2}\right). \tag{12}$$

Hence

$$\mathbb{E}[S] = \nu e^{\mu+\sigma^2/2}, \quad \text{Var}(S) = \nu e^{2\mu+2\sigma^2}.$$

The squared coefficient of variation is

$$\text{CV}_S^2 = \frac{\text{Var}(S)}{\mathbb{E}[S]^2} = \frac{e^{\sigma^2}}{\nu}. \tag{13}$$

Skewness is

$$\gamma_{1,S} = \frac{\mu_3(S)}{\text{Var}(S)^{3/2}} = \frac{e^{3\sigma^2/2}}{\sqrt{\nu}}. \tag{14}$$

Equations (13) and (14) illustrate how inequality arises from two channels: participation/earner counts (ν) and wage dispersion (σ^2).

Remark 5.1 (Truncation adjustment). If using $M \mid N \geq 1$, $\mathbb{P}(S = 0)$ changes, and unconditional moments become conditional-on-truncation moments. In empirical work, parameters are estimated directly from observed data (Section 9) rather than relying on closed-form corrections.

Remark 5.2 (Laplace transform). For $t \geq 0$, the Laplace transform of a compound Poisson sum is

$$\mathcal{L}_S(t) = \mathbb{E}[e^{-tS}] = \exp\{\nu(\mathcal{L}_Z(t) - 1)\},$$

but for lognormal Z, $\mathcal{L}_Z(t)$ has no elementary closed form and is evaluated numerically.

6 Inequality Measurement

The presence of a structural mass at zero income fundamentally shapes inequality in our model. Standard inequality indices, such as the Gini coefficient, must account for both the discrete probability of zero earnings and the continuous dispersion among positive earners. Fortunately, the Gini index admits a clean decomposition in this setting.

This decomposition is especially useful from a policy perspective because it separates two conceptually different sources of disadvantage: exclusion from labor income altogether and unequal earnings among those who do participate. In applied work, these channels often imply different policy responses, for example, employment-support measures for the former and wage-compression or transfer policies for the latter.

Let G_S denote the Gini coefficient of total labor income S. Because S is a mixture of a point mass at zero and a continuous distribution over $(0, \infty)$, we can write:

$$G_S = \mathbb{P}(M = 0) + \left[1 - \mathbb{P}(M = 0)\right] G_{S|M>0}, \tag{15}$$

where:

- $\mathbb{P}(M = 0) = e^{-\nu}$ (under the untruncated Poisson assumption) is the share of households with no earners;
- $G_{S|M>0}$ is the Gini coefficient computed only over households with positive labor income.

This decomposition follows from the general formula for the Gini coefficient of a distribution with a mass at zero (see, e.g., Cowell [9]). Intuitively, every zero-income household is "maximally unequal" relative to any positive-income household, contributing directly to overall inequality. In fact, $\mathbb{P}(M = 0)$ serves as a *lower bound* on G_S: even if all working households earned identical incomes ($G_{S|M>0} = 0$), inequality would still be positive due to non-participation.

Two key forces drive total inequality:

1. **Labor force participation** (p): Lower participation increases $\mathbb{P}(M = 0)$, raising the first term. This channel dominates in contexts with high unemployment, early retirement, or low female labor supply.
2. **Wage dispersion** (σ^2): Higher heterogeneity in log-wages inflates $G_{S|M>0}$, especially since income sums inherit and amplify skewness (as shown in Section 5). This channel is central to rising top-end inequality.

Notably, household size λ affects inequality through both channels: larger λ raises expected earners ($\nu = \lambda p$), reducing $\mathbb{P}(M = 0)$, while also diversifying wage risk and lowering $G_{S|M>0}$ (via the $1/\sqrt{\nu}$ scaling of skewness). Thus, demographic structure and labor market institutions jointly determine the inequality landscape.

This interpretation gives the decomposition practical value beyond formal measurement. If inequality is driven mainly by the extensive margin, then policies that raise participation or stabilize job access may be more effective than policies aimed only at wage dispersion. If, instead, most inequality remains within positive-earning households, then the main levers lie in wage-setting, transfers, or tax–benefit design.

When using truncated data ($N \geq 1$), the decomposition becomes:

$$G_S = \mathbb{P}(M = 0 \mid N \geq 1) + \left[1 - \mathbb{P}(M = 0 \mid N \geq 1)\right] G_{S|M>0,N\geq 1},$$

with $\mathbb{P}(M = 0 \mid N \geq 1) = (e^{-\nu} - e^{-\lambda})/(1 - e^{-\lambda})$ as derived earlier. Ignoring this adjustment leads to underestimation of the zero-earner share and thus of overall inequality.

7 Approximating Sums of Lognormals

Two practical approaches are employed:

(i) Fenton–Wilkinson (FW) lognormal approximation. For fixed m, $S_m = \sum_{k=1}^{m} Z_k$ is approximated by $\tilde{S}_m \sim \text{Lognormal}(\mu_m, \sigma_m^2)$ matching the first two moments:

$$\sigma_m^2 = \log\left(1 + \frac{e^{\sigma^2} - 1}{m}\right), \quad \mu_m = \log m + \mu + \frac{\sigma^2 - \sigma_m^2}{2}. \tag{16}$$

FW is simple and often accurate in the right tail.

(ii) Monte Carlo evaluation. For microsimulation and policy analysis, Monte Carlo is straightforward and can incorporate covariates, survey-calibrated parameters, and non-labor components R.

8 Consumption: Conditional and Unconditional Properties

From the Engel curve, conditional on resources $Y = y$,

$$\mathbb{E}[C \mid Y = y] = \exp\left(\beta_0 + \frac{\tau^2}{2}\right)(y + \delta)^{\beta_1}. \tag{17}$$

The expected marginal propensity to consume is

$$\mathrm{MPC}(y) = \frac{d}{dy}\mathbb{E}[C \mid Y = y] = \beta_1 \exp\left(\beta_0 + \frac{\tau^2}{2}\right)(y+\delta)^{\beta_1 - 1}, \tag{18}$$

which declines in y when $\beta_1 < 1$. This concavity amplifies inequality reduction from equal transfers.

Unconditionally,

$$\mathbb{E}[C] = \exp\left(\beta_0 + \frac{\tau^2}{2}\right)\mathbb{E}[(Y+\delta)^{\beta_1}], \tag{19}$$

evaluated by Monte Carlo or by combining FW for S with a parametric model for R.

Figure 1 displays simulated inequality outcomes under alternative UBI scenarios. It compares baseline and post-transfer distributions for household resources and consumption and summarizes the associated inequality changes. The figure shows nonlinear Gini reductions in both income and consumption, with stronger compression in consumption because Engel-curve concavity implies proportionally larger welfare gains for lower-resource households.

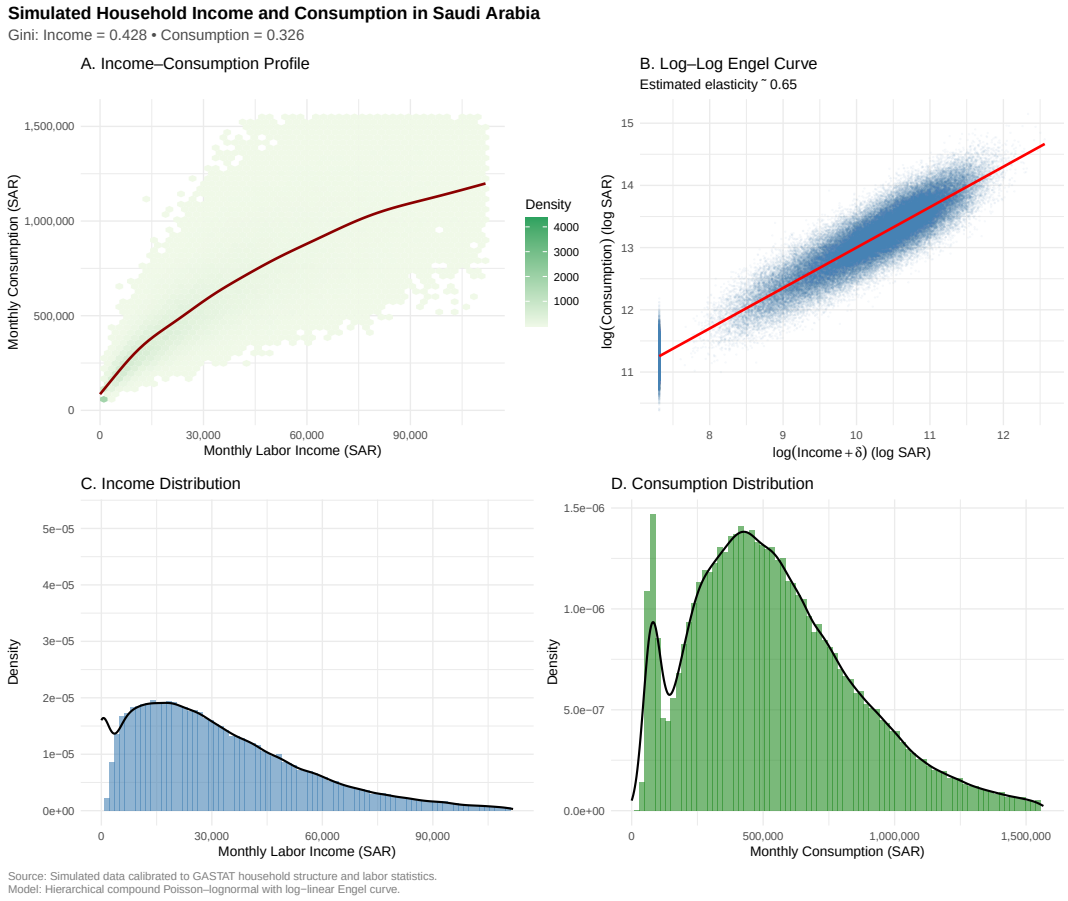

Figure 1: Microsimulation results for 100,000 synthetic households calibrated to Saudi HICES 2023 aggregates. The figure compares baseline and post-UBI distributions of household resources and consumption and illustrates that a uniform transfer compresses the lower tail of the distribution, with a larger reduction in consumption inequality due to the concavity of the Engel curve.

As shown in Figure 1, the model reproduces realistic baseline dispersion patterns while also clarifying the mechanism through which the policy shock operates. The reduction in lower-tail deprivation is immediate in the income distribution because the transfer raises the resources of all households by the same absolute amount, including those with zero or very low labor income. The corresponding reduction in consumption inequality is stronger because the estimated Engel specification is concave, so identical

increments in resources translate into proportionally larger consumption gains for poorer households. This visual evidence is consistent with the model's central interpretation that participation and wage heterogeneity drive labor-income inequality, whereas nonlinear consumption behavior amplifies the equalizing effect of transfers on welfare outcomes.

9 Estimation and Calibration with Saudi HICES

This section provides a design-consistent workflow for HICES microdata.

9.1 Constructing Model Objects from HICES Variables

Using household and person records, define the following variables, aligned to monthly units:

- Household size N_i: number of household members (roster count).
- Earners M_i: members with positive labor earnings over the income reference period.
- Individual labor earnings Z_{ik}: wage/salary and self-employment labor components.
- Household labor income $S_i = \sum_k Z_{ik}$.
- Non-labor resources R_i: remittances, transfers, pensions, and property income.
- Consumption C_i: monthly aggregate constructed using HICES category-specific reference periods, GASTAT [2].
- Survey weight w_i.

HICES reference periods matter: income is attributed to the prior 12 months, while diary spending covers the subsequent 14 days, and other categories use 1-, 3-, or 12-month recall windows, GASTAT [2]. Constructing C_i should follow the official aggregation protocol.

9.2 Participation Model

Rather than a constant p, use a covariate model:

$$M_i \mid N_i \sim \text{Binomial}(N_i, p_i), \quad \text{logit}(p_i) = \mathbf{x}_i^\top \boldsymbol{\alpha},$$

estimated via weighted binomial GLM.

9.3 Wage Model

On the worker subsample,

$$\log Z_{ik} = \mathbf{u}_{ik}^\top \boldsymbol{\beta} + \eta_{ik}, \quad \eta_{ik} \sim \mathcal{N}(0, \sigma^2),$$

estimated via weighted least squares.

9.4 Engel Curve Estimation

Estimate via weighted regression:

$$\log C_i = \beta_0 + \beta_1 \log(Y_i + \delta) + \mathbf{q}_i^\top \boldsymbol{\gamma} + \varepsilon_i,$$

where $\mathbf{q}_i$ may include household size and region controls. Choose δ as a small constant or estimate by nonlinear least squares.

9.5 Validation Against Published Aggregates

As external validity checks, reproduce published monthly means for disposable income and consumption expenditure from the 2023 release, GASTAT [13].

9.6 Illustrative Empirical Application

Data are simulated for 100,000 households using parameters calibrated to approximate Saudi aggregates (average disposable income ≈ 11,839 SAR, consumption ≈ 10,884 SAR, GASTAT [13]). True parameters: $\lambda = 4.0$, $p = 0.4$, wage lognormal with $\mu = 8.5$, $\sigma = 0.8$, R lognormal yielding average $Y \approx 11{,}861$ SAR, Engel $\beta_0 = 1.0$, $\beta_1 = 0.85$, $\delta = 100$, $\tau = 0.3$.

Estimated parameters recover the true values closely (Table 1).

Table 1: Estimated parameters from simulated data.

Parameter	Estimate
λ (household size)	3.99
p (participation prob.)	0.41
μ (wage log-mean)	8.49
σ (wage log-sd)	0.80
μ_R (non-labor log-mean)	6.00
σ_R (non-labor log-sd)	1.04
β_0 (Engel intercept)	1.00
β_1 (Engel slope)	0.85
τ (Engel residual sd)	0.32

Inequality diagnostics: jobless share $p_0 = 0.170$, Gini for positive $S = 0.439$, overall Gini for $Y = 0.507$, and Gini for $C = 0.475$.

Relative to existing flexible heavy-tailed and truncated models, these results demonstrate the practical advantage of the proposed framework. While earlier studies such as Hassan et al. [3] and Ahmadini et al. [5] show that generalized and truncated distributions can improve fit for skewed data, they do not jointly model structural zeros, stochastic earner counts, and downstream consumption behavior. Likewise, Lomax-based and generalized Beta extensions such as Kamal et al. [4] and Mahmoud et al. [6] provide useful flexible parametric forms, but they are not designed for household-level microsimulation under survey constraints. In contrast, the present model not only recovers the true parameters closely in simulation, but also produces interpretable inequality diagnostics and supports policy counterfactuals within a unified framework. The stronger inequality compression observed for consumption than for income under the UBI exercise further highlights a novel feature of the approach: distributional effects can be traced simultaneously to participation, wage dispersion, and nonlinear Engel-curve transmission.

10 Policy Application: Universal Basic Income

Consider a monthly UBI transfer $G > 0$ per household. Post-transfer resources are

$$Y_i' = Y_i + G, \quad \log C_i' = \beta_0 + \beta_1 \log(Y_i' + \delta) + \varepsilon_i.$$

With $\beta_1 < 1$, Engel concavity compresses consumption more than income.

In the simulation, $G = 1{,}000$ SAR reduces Gini Y from 0.507 to 0.467, and Gini C from 0.475 to 0.438.

The policy interpretation is straightforward. A uniform transfer improves the position of all households in absolute terms, but its relative impact is largest for households located at the bottom of the resource distribution. Because consumption responds concavely to resources, the same transfer induces a stronger equalizing effect in welfare-relevant consumption than in pre-consumption resources themselves. This result helps explain why policies that appear modest when judged only by income inequality can still generate more substantial reductions in consumption-based inequality and material deprivation.

Saudi calibration anchors. The 2023 publication reports average monthly disposable income of 11,839 SAR and consumption expenditure of 10,884 SAR, GASTAT [13]. These serve as targets for calibrating β_0 and verifying the aggregates Y and C.

11 Limitations

This study has several limitations. First, the baseline specification assumes independence across core components unless otherwise stated, whereas household size, participation decisions, wages, and non-labor resources may be correlated through unobserved socioeconomic factors. Second, wages are modeled as lognormal and participation as conditionally Bernoulli, choices that support analytical tractability but may not fully capture labor-market segmentation or extremely heavy upper-tail outcomes. Third, the empirical illustration is based on calibrated simulation rather than confidential scientific-use Saudi HICES microdata, so the numerical results should be interpreted as methodological validation rather than definitive empirical estimates. Fourth, the consumption equation is static and reduced-form and, therefore, does not explicitly model intertemporal smoothing, behavioral feedback, or labor-supply responses to transfers. Finally, the UBI exercise is partial-equilibrium and does not incorporate financing mechanisms, price responses, or broader general-equilibrium adjustments.

12 Conclusion

This study develops a hierarchical probabilistic framework for household labor income and consumption that is specifically designed for microsimulation and policy analysis. Labor income is represented as a compound count–lognormal process generated by zero-truncated household size, Bernoulli labor participation, and lognormal wages, thereby capturing both the structural mass at zero and the heavy right tail observed in household income data. Consumption is linked to total resources through a log-linear Engel curve that accommodates concavity and heteroskedasticity in spending behavior. A central contribution of the study is that the framework remains analytically tractable, yielding closed-form cumulants and moments, supporting lognormal-sum approximations, and accommodating survey-weighted estimation consistent with Saudi HICES reference periods and design features. The simulation results show accurate parameter recovery and realistic inequality diagnostics, while the UBI counterfactual demonstrates that equal cash transfers reduce inequality nonlinearly, with stronger effects for consumption than for income because of Engel-curve concavity. These findings indicate that the proposed framework is not only statistically rigorous but also substantively useful for evaluating welfare and distributional policies. Future research may extend the model by allowing richer dependence across household members, incorporating alternative household-size distributions and explicit non-labor income processes, estimating the framework using full Saudi HICES microdata, and introducing dynamic behavioral responses together with general-equilibrium channels for more comprehensive policy evaluation.

Acknowledgments

The author thanks the Editor and anonymous referees for constructive feedback. All remaining errors are the author's responsibility.

Funding Statement

No funding was received for this study.

Data Availability

The data set presented in this article is not readily available. Requests for access to the dataset should be directed to the corresponding author.

Conflicts of Interest:

There are no conflicts of interest with respect to this study.

Author Contributions

The author is solely responsible for all components of this work.

AI Usage Statement:

No AI-based tools were used in the generation, writing, revising, analysis, or interpretation of the research data and findings presented in this article.

Appendix

Algorithms (Survey-Weighted)

Algorithm 1: Weighted Gini coefficient for nonnegative data (zeros allowed)

Input: Values $x_i \geq 0$ and sampling weights $w_i > 0$, $i = 1, \ldots, n$.
Output: Weighted Gini coefficient $G_w \in [0, 1]$.
Remove missing/non-finite values from (x_i, w_i) and keep only $w_i > 0$ and $x_i \geq 0$;
Normalize weights: $w_i \leftarrow w_i / \sum_j w_j$;
Sort pairs (x_i, w_i) by x_i ascending;
Compute weighted mean: $\mu \leftarrow \sum_i w_i x_i$;
if $\mu = 0$ **then**
  return 0;
Compute cumulative weights $W_i = \sum_{j \leq i} w_j$;
Compute cumulative weighted shares $L_i = \frac{1}{\mu} \sum_{j \leq i} w_j x_j$;
Approximate Lorenz area by trapezoids: $A \leftarrow \sum_{i=1}^{n-1} (W_{i+1} - W_i) \cdot \frac{L_{i+1} + L_i}{2}$;
return $G_w \leftarrow 1 - 2A$;

Simulation on HICES-style Units

In HICES applications, simulate counterfactuals by reweighting or perturbing observed microdata to preserve joint distributions of demographics and baseline outcomes. A design-consistent approach is:

- Start from observed household resources Y_i and consumption C_i.
- Define $Y_i' = Y_i + G$ for a UBI G.
- Predict C_i' by the Engel curve: draw ε_i or use conditional expectation.
- Compute weighted inequality indices on (Y_i', C_i') using Algorithm 1.

References

[1] Fenton, L. (1960). The sum of log-normal probability distributions in scatter transmission systems. IRE Transactions on Communications Systems, 8(1), 57-67. doi:10.1109/TCOM.1960.1097606.

[2] General Authority for Statistics (GASTAT) (2024). *Methodology and Quality Report of Household Income and Consumption Expenditure Statistics (HICES)*, Riyadh: GASTAT.

[3] Hassan, A. S., Hemeda, S. E., & Nassr, S. G. (2021). On the extension of exponentiated Pareto distribution. Journal of Modern Applied Statistical Methods, 19(1), 21. doi:10.22237/jmasm/1619481840.

[4] Kamal, M., Aldallal, R., Nassr, S. G., Al Mutairi, A., Yusuf, M., Mustafa, M. S., Alsolmi, M. M., & Almetwally, E. M. (2023). A new improved form of the Lomax model: Its bivariate extension and an application in the financial sector. Alexandria Engineering Journal, 75, 127-138. doi:10.1016/j.aej.2023.05.027.

[5] Ahmadini, A. A. H., Hassan, A., Elgarhy, M., Elsehetry, M., Alshqaq, S. S., & Nassr, S. G. (2021). Inference of truncated Lomax inverse Lomax distribution with applications. Intelligent Automation & Soft Computing, 29(1), 199-212. doi:10.32604/iasc.2021.017890.

[6] Mahmoud, A. A. M., Seyam, E. A., Khashab, R. H., Alshawarbeh, E., Kalantan, Z. I., Khalil, O. F., El-Saeed, A. R., & Nassr, S. G. (2025). A New extension of generalized beta distribution: statistical properties, prediction estimation and application in failure components and physics data. Scientific African, e03037. doi:10.1016/j.sciaf.2025.e03037.

[7] Tang, X., Seong, J. T., Alharbi, R., Al Mutairi, A., & Nasr, S. G. (2024). A new probabilistic model: Theory, simulation and applications to sports and failure times data. Heliyon, 10(4). doi:10.1016/j.heliyon.2024.e25651.

[8] Kleiber, C., & Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*, New York: Wiley.

[9] Cowell, F. A. (2011). *Measuring Inequality*, 3rd ed., Oxford: Oxford University Press.

[10] Atkinson, A. B. (1970). On the measurement of inequality. Journal of Economic Theory, 2, 244–263. `https://faculty.ucr.edu/~jorgea/econ261/atkinson_inequality.pdf`.

[11] Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Washington, DC: World Bank.

[12] Willmot, G. E., & Lin, X. S. (2001). *Lundberg Approximations for Compound Distributions with Insurance Applications*, New York: Springer.

[13] General Authority for Statistics (GASTAT) (2023). *Household Income and Consumption Expenditure Survey Publication 2023*, Riyadh: GASTAT.

Disclaimer / Publisher's Note:

The statements and data in all publications are solely those of the author(s) and contributor(s) and do not necessarily reflect the views of the publisher. The publisher accepts no responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.